Cointegrating Polynomial Regressions with Power Law Trends

Lin, Yicong; Reuvers, Hanno

2021

Cointegrating Polynomial Regressions with Power Law Trends: A New Angle on the Environmental Kuznets Curve*

Yicong Lin†[1] and Hanno Reuvers[2]

[1]Department of Econometrics and Data Science, Vrije Universiteit Amsterdam, 1081 HV Amsterdam, The Netherlands

[2]Department of Econometrics, Erasmus University Rotterdam, 3062 PA Rotterdam, The Netherlands

September 7, 2020

Abstract

The Environment Kuznets Curve (EKC) predicts an inverted U-shaped relationship between economic growth and environmental pollution. Current analyses frequently employ models which restrict the nonlinearities in the data to be explained by the economic growth variable only. We propose a Generalized Cointegrating Polynomial Regression (GCPR) with flexible time trends to proxy time effects such as technological progress and/or environmental awareness. More specifically, a GCPR includes flexible powers of deterministic trends and integer powers of stochastic trends. We estimate the GCPR by nonlinear least squares and derive its asymptotic distribution. Endogeneity of the regressors can introduce nuisance parameters into this limiting distribution but a simulated approach nevertheless enables us to conduct valid inference. Moreover, a subsampling KPSS test can be used to check the stationarity of the errors. A comprehensive simulation study shows good performance of the simulated inference approach and the subsampling KPSS test. We illustrate the GCPR approach on a dataset of 18 industrialised countries containing GDP and CO_2 emissions. We conclude that: (1) the evidence for an EKC is significantly reduced when a nonlinear time trend is included, and (2) a linear cointegrating relation between GDP and CO_2 around a power law trend also provides an accurate description of the data.

JEL Classification: C12, C13, C32, O44, Q20

Keywords: Cointegration Testing, Environmental Kuznets Curve, Generalized Cointegrating Polynomial Regression, Nonlinear Least Squares, Power Law Trends

*Earlier versions of this paper have been presented at the Econometrics Internal Seminar (EIS) at Erasmus University Rotterdam, and the 2019 CFE meeting in London. We gratefully acknowledge the comments by the participants. We extend our thanks to Eric Beutner, Dick van Dijk, and Stephan Smeekes for their valuable feedback.

†Corresponding author: yc.lin@vu.nl.

1 Introduction

On page 370 of their seminal paper, Grossman and Krueger (1995) conclude:

> "*Contrary to the alarmist cries of some environmental groups, we find no evidence that economic growth does unavoidable harm to the natural habitat. Instead we find that while increases in GDP may be associated with worsening environmental conditions in very poor countries, air and water quality appear to benefit from economic growth once some critical level of income has been reached.*"

The quote above suggests an inverted U-shaped relationship between environmental degradation and economic growth. This relationship is currently known as the Environmental Kuznets Curve (EKC) and it forms an active research area. Indeed, some 25 years after its first conception, there now exists a rich literature that (1) reports on the experimental evidence on the existence/nonexistance of the EKC, (2) provides economic theory to explain the EKC, and/or (3) refines the econometric tools that are used to analyse the EKC.[1] The Web of Science returns a list of over 2,900 articles when the search query "*Environmental Kuznets Curve*" is entered.[2]

The studies on the EKC have been criticised on two main points. First, the GDP variable was initially treated as a stationary variable even though unit root tests do not reject the null hypothesis of a unit root. This has further consequences since EKC regressions include higher integer powers of GDP as well. The combination of nonstationarity and nonlinearity places the EKC in the nonlinear cointegration literature and appropriate econometric techniques should be employed. Such techniques have been provided in Wagner (2015) and Wagner and Hong (2016) under the name of *Cointegrating Polynomial Regressions* (CPRs), that is regressions containing: deterministic variables, integrated processes, and integer powers of integrated processes.[3] These CPRs are estimated by fully modified OLS to allow for standard inference on the coefficients. Model diagnostics and a multiple countries analysis are discussed in Wang et al. (2018) and Wagner et al. (2019), respectively.

Second, there is an ongoing debate in the EKC literature on the model specification, and more specifically, on omitted variables. Omitted variables are a valid concern because adaptation to clean technology,[4] pollution control policy,[5] increasing energy efficiency, and increasing environmental awareness may all influence pollution levels yet are also difficult quantify (and for that reason often excluded from the reduced-form model). It has been argued that the inclusion of deterministic time trends will control for such omitted variables. In empirical applications, this typically translates

[1]Further references to these specific areas of research can be found in the review articles by Dasgupta et al. (2002), Stern (2004), and Carson (2009) among others.

[2]Web of Science, accessed on August 27, 2020, http://www.webofknowledge.com.

[3]The recent work by Stypka et al. (2017) confirms the importance of treating the growth variable as nonstationary. However, it appears less important to use an estimation procedure that incorporates the fact that several integer powers of the same integrated process appear as regressors. Namely, Stypka et al. (2017) also find that the "standard estimator" which treats higher order powers of the integrated regressor as additional I(1) variables has the same limiting distribution as the CPR estimator (yet a slightly worse finite sample performance).

[4]Nordhaus (2014) discusses the link between climate change and technological changes. As another example, Figure 2 in Gillingham and Stock (2018) reports a steady decline in the price of solar panels and a steady growth in solar panel sales. Cheaper solar energy can substitute fossil energy thereby reducing pollution.

[5]A policy variable, 'Repudiation of Contracts by Government', was included by Panayotou (1997) to proxy the quality of environmental policies and institutions.

into the inclusion of a linear deterministic trend, see Panayotou (1997) and Stypka et al. (2017), for example.

There seems no a priori reason why *linear* deterministic trends should control for omitted variables and provide a valid EKC specification. On the contrary, we will reason here, and later also in the empirical application, that omitted nonlinear trends are more likely to result in erroneous EKC results. The small simulation setting in Table 1 illustrates this point. We consider an omitted nonlinear deterministic trend and estimate an EKC type of regression: $y_t = \tau_1 + \tau_2 t + \phi_1 x_t + \phi_2 x_t^2 + u_t$. We test $H_0 : \phi_2 \geq 0$ versus $H_a : \phi_2 < 0$ because a significantly negative coefficient in front of x_t^2 is the typical result economists associate with the existence of the EKC. It is seen how a (correctly sized) Wald test misinterprets the negative curvature of a deterministic trend for negative curvature caused by a squared integrated variable. In other words, a negative and significant coefficient in front of the square of GDP might be caused by omitted nonlinear deterministic trends rather than being any evidence for the EKC.

The current paper augments the Cointegrating Polynomial Regressions of Wagner and Hong (2016) with power law deterministic trends. The powers of these time trends are estimated and thereby provide additional flexibility in explaining the nonlinear trending behaviour observed in the data. We provide the limiting distribution of the estimator and propose a simulated approach for parameter inference. Additionally, we show how a KPSS-type of test remains useful in verifying the stationarity of the error process hence avoiding spurious results or misspecification of the cointegrating relation. A Monte Carlo study sheds light on the finite sample properties of the simulated approach and stationarity test. As an empirical application, we revisit a dataset on 18 countries over the timespan 1870-2014 and study the Environmental Kuznets Curve. For each of these countries, we find that the flexible deterministic trends sufficiently capture the nonlinearities in the data and turn higher integer powers of log per capita GDP redundant.

Our paper builds upon several different strands of literature. Clearly, we rely on results from the literature on Cointegrating Polynomial Regressions (see references above). Additionally, various references on power law trends are closely related to the current work. Phillips (2007) provides a detailed analysis of the power law trend regression. An extension of such power law regressions to spatial lattices is covered in Robinson (2012). Finally, there is also recent work by Hu et al. (2019) on power law functions applied to stochastic trends.

This paper is organized as follows. Section 2 introduces the model and the estimation framework. Asymptotic properties of the estimators and parameter inference are discussed in Section 3. The Monte Carlo simulations in Section 4 compare asymptotic results and the finite sample distributions. An in depth discussion of the Environment Kuznets Curve can be found in Section 5. Section 6 concludes. All proofs can be found in the Appendix.

Finally, some words on notation. The integer part of the number $a \in \mathbb{R}^+$ is denoted by $[a]$. For a vector $\boldsymbol{x} \in \mathbb{R}^n$, its p-norm is denoted by $\|\boldsymbol{x}\|_p = (\sum_{i=1}^n |x_i|^p)^{1/p}$. For a matrix $\boldsymbol{A}$, say of dimension $(n \times m)$, the induced p-norm and Frobenius norm are defined as $\|\boldsymbol{A}\|_p = \sup_{\boldsymbol{x} \neq \boldsymbol{0}} \|\boldsymbol{A}\boldsymbol{x}\|_p / \|\boldsymbol{x}\|_p$ and $\|\boldsymbol{A}\|_{\mathcal{F}} = \sqrt{\sum_{i=1}^n \sum_{j=1}^m a_{ij}^2}$, respectively. For p-norms, we will omit the subscripts whenever $p = 2$. $\boldsymbol{I}_n$ denotes the $(n \times n)$ identity matrix. Two special linear algebra operators are: the Hadamard product (element-wise multiplication) denoted by "$\odot$" and the Kronecker product denoted by "$\otimes$". We omit the integration bounds whenever the integral is take over $[0, 1]$. The symbol "$\Rightarrow$" signifies weak convergence, "$\overset{d}{=}$" stands for equality in distribution, and "$\longrightarrow_p$" and "$\longrightarrow_d$" denote convergence in probability and in distribution. If convergence occurs conditionally on the sample, then we add a superscript "*" to the standard notation. The probabilistic Landau symbols are $O_p(\cdot)$ and $o_p(\cdot)$. Finally, the generic constant C can change from line to line.

2 The Model and NLS Estimation

Our model specification is a hybrid of the power law regressions from Robinson (2012) and the cointegrating polynomial regression (CPR) introduced by Wagner and Hong (2016). It combines integrated regressors (and their integer powers) with a flexible deterministic trend specification. The resulting *Generalized Cointegrating Polynomial Regression* (GCPR) for y_t is given by

$$y_t = \sum_{i=1}^{d} \tau_i t^{\theta_i} + \sum_{i=1}^{m} \sum_{j=1}^{p_i} \phi_{ij} x_{it}^j + u_t = \boldsymbol{d}_t(\boldsymbol{\theta})'\boldsymbol{\tau} + \sum_{i=1}^{m} \boldsymbol{x}'_{(i)t}\boldsymbol{\phi}_i + u_t = \boldsymbol{d}_t(\boldsymbol{\theta})'\boldsymbol{\tau} + \boldsymbol{s}'_t\boldsymbol{\phi} + u_t, \tag{2.1}$$

where $\boldsymbol{\theta} = [\theta_1, \theta_2, \ldots, \theta_d]'$, $\boldsymbol{\tau} = [\tau_1, \tau_2, \ldots, \tau_d]'$, $\boldsymbol{\phi}_i = [\phi_{i1}, \phi_{i2}, \ldots, \phi_{i,p_i}]'$, $\boldsymbol{d}_t(\boldsymbol{\theta}) = [t^{\theta_1}, \ldots, t^{\theta_d}]'$ and $\boldsymbol{x}_{(i)t} = [x_{it}, x_{it}^2, \ldots, x_{it}^{p_i}]'$ collects the integer powers of the i^{th} integrated regressor ($i = 1, 2, \ldots, m$). The final equality in (2.1) relies on the definitions $\boldsymbol{s}_t = [\boldsymbol{x}'_{(1)t}, \boldsymbol{x}'_{(2)t}, \ldots, \boldsymbol{x}'_{(m)t}]'$ and $\boldsymbol{\phi} = [\boldsymbol{\phi}'_1, \boldsymbol{\phi}'_2, \ldots, \boldsymbol{\phi}'_m]'$. The error term u_t is stationary (see Assumption 2 for more details).

We consider nonlinear least squares (NLS) estimators of the unknown parameters in (2.1). As such, we define the objective function $Q_T(\boldsymbol{\theta}, \boldsymbol{\tau}, \boldsymbol{\phi}) = \frac{1}{2}\sum_{t=1}^{T}\left(y_t - \boldsymbol{d}_t(\boldsymbol{\theta})'\boldsymbol{\tau} - \boldsymbol{s}'_t\boldsymbol{\phi}\right)^2$ and compute

$$\left(\widehat{\boldsymbol{\theta}}_T, \widehat{\boldsymbol{\tau}}_T, \widehat{\boldsymbol{\phi}}_T\right) = \underset{(\boldsymbol{\theta},\boldsymbol{\tau},\boldsymbol{\phi})\in\Theta\times\mathbb{R}^d\times\mathbb{R}^p}{\arg\min} Q_T(\boldsymbol{\theta}, \boldsymbol{\tau}, \boldsymbol{\phi}), \tag{2.2}$$

where $p = \sum_{i=1}^{m} p_i$,

$$\Theta = \left\{\theta_1, \theta_2, \ldots, \theta_d : -\tfrac{1}{2} < \theta_L \leq \theta_1;\ \theta_j - \theta_{j-1} \geq \delta, j = 2, \ldots, d;\ \theta_d \leq \theta_U < \infty\right\} \subset \mathbb{R}^p,$$

for some lower bound θ_L, some upper bound θ_U, and $\delta > 0$. Note that Θ is closed and bounded and therefore compact.

The optimization problem in (2.2) is easy to solve. For any given $\boldsymbol{\theta} \in \Theta$, the minimizers for $\boldsymbol{\tau}$ and $\boldsymbol{\phi}$ can be found from the OLS regression

$$\begin{bmatrix} \boldsymbol{\tau}(\boldsymbol{\theta}) \\ \boldsymbol{\phi}(\boldsymbol{\theta}) \end{bmatrix} = \left(\sum_{t=1}^{T} \boldsymbol{z}_t(\boldsymbol{\theta})\boldsymbol{z}_t(\boldsymbol{\theta})'\right)^{-1} \left(\sum_{t=1}^{T} \boldsymbol{z}_t(\boldsymbol{\theta})y_t\right), \text{ with } \boldsymbol{z}_t(\boldsymbol{\theta}) = [\boldsymbol{d}_t(\boldsymbol{\theta})', \boldsymbol{s}'_t]'. \tag{2.3}$$

We can thus minimize the concentrated criterion function $\widetilde{Q}_T(\boldsymbol{\theta}) = Q_T\big(\boldsymbol{\theta}, \boldsymbol{\tau}(\boldsymbol{\theta}), \boldsymbol{\phi}(\boldsymbol{\theta})\big)$ to obtain $\widehat{\boldsymbol{\theta}}_T$ and run a subsequent OLS regression to find $\widehat{\boldsymbol{\tau}}_T$ and $\widehat{\boldsymbol{\phi}}_T$.

Remark 1
The fixed powers of x_{it} allow us to test for their significance and thereby distinguish between nonlinearities caused by deterministic and/or stochastic trends. This is important for our empirical application on the Environmental Kuznets Curve, see Section 5. Hu et al. (2019) study a model with a flexible power of the integrated regressor. That is, these authors derive the limiting distribution of the NLS estimators for β and γ when $y_t = \beta|x_t|^\gamma + u_t$ with $\beta \neq 0$.

Remark 2
Equations (2.1)-(2.2) *assume that each of the d elements in $\boldsymbol{\theta}$ needs to be estimated. First, we*

envision model specifications where d is small such that the deterministic trends cannot represent the integrated regressors (see, e.g. Phillips (1998)). Second, the applied researcher might prefer to fix certain elements of this vector. For example, requiring $\theta_1 = 0$ *and* $\theta_2 = 1$ *includes an intercept and linear trend into the model. A typical model specification could be*

$$y_t = \tau_1 + \tau_2 t + \tau_3 t^{\theta} + \boldsymbol{s}_t'\boldsymbol{\phi} + u_t. \tag{2.4}$$

Our approach and proofs are easily adapted to the case in which some elements of $\boldsymbol{\theta}$ *are prespecified.*

3 Asymptotic Theory

We subsequently study the asymptotic properties of the NLS estimators. To this end we first collect all the unknown parameters in the vector $\boldsymbol{\gamma} = [\boldsymbol{\theta}', \boldsymbol{\tau}', \boldsymbol{\phi}']'$. This vector is assumed to be an element of the parameter space $\boldsymbol{\Gamma} = \boldsymbol{\Theta} \times \mathbb{R}^{d+p}$. The true parameter vector is $\boldsymbol{\gamma}_0 = [\boldsymbol{\theta}_0', \boldsymbol{\tau}_0', \boldsymbol{\phi}_0']'$.

Assumption 1
For all $1 \le i \le d$, *we have* $\tau_{0i} \neq 0$.

Assumption 2
Let $\boldsymbol{\zeta}_t = [\eta_t', \boldsymbol{\varepsilon}_t']'$ *be a sequence of i.i.d. random vectors with* $\mathbb{E}(\boldsymbol{\zeta}_t) = \mathbf{0}$, $\boldsymbol{\Sigma} = \mathbb{E}(\boldsymbol{\zeta}_t\boldsymbol{\zeta}_t')$, *and* $\mathbb{E}\,\|\boldsymbol{\zeta}_t\|^{2q} < \infty$ *for some* $q > 2$.

(a) $u_t = \sum_{k=0}^{\infty} \psi_k \eta_{t-k}$ *with* $\sum_{k=1}^{\infty} k|\psi_k| < \infty$.

(b) $\boldsymbol{x}_t = \sum_{s=1}^{t} \boldsymbol{v}_s$, *where* $\boldsymbol{v}_t = \sum_{k=0}^{\infty} \boldsymbol{\Psi}_k \boldsymbol{\varepsilon}_{t-k}$ *with* $\sum_{k=0}^{\infty} \|\boldsymbol{\Psi}_k\| < \infty$ *and* $\det\left(\sum_{k=0}^{\infty} \boldsymbol{\Psi}_k\right) \neq 0$.

The first assumption is needed to avoid identification issues. That is, if $\tau_{0i} = 0$ for some $i \in \{1, \ldots, d\}$, then the corresponding θ_{0i} is not identified and the Davies problem arrises when testing $H_0 : \tau_i = 0$ (see Davies (1977, 1987)). We will not consider such difficulties in the current paper and this is reflected in our model specification (2.1). That is, we consider *flexible* powers of the deterministic trends but *fixed* powers of the stochastic trends thus allowing us to test zero restrictions on (elements of) $\boldsymbol{\phi}$. This is of crucial importance in the EKC application to see whether nonlinear effects in the economic growth variable remain significant after nonlinear time trends have been added to the model. For different model settings Assumption 1 has been relaxed in the literature. Baek et al. (2015) and Cho and Phillips (2018) study the asymptotic behaviour of a quasi-likelihood ratio test when Assumption 1 is violated and the conditional mean of the data contains strictly stationary regressors and a flexible time trend. Alternatively, one can use drifting parameter sequences with different identification strengths as in Andrews and Cheng (2012).

Assumption 2 excludes cointegration among elements of $\boldsymbol{x}_t$ and defines it as the partial sum of a short memory process. The latter implies that

$$\frac{1}{\sqrt{T}} \sum_{t=1}^{[rT]} \begin{bmatrix} u_t \\ \boldsymbol{v}_t \end{bmatrix} \Rightarrow \boldsymbol{B}(r) = \begin{bmatrix} B_u(r) \\ \boldsymbol{B}_v(r) \end{bmatrix} \tag{3.1}$$

where $\boldsymbol{B}(r)$ denotes an $(m+1)$-dimensional vector Brownian motion with covariance matrix $\boldsymbol{\Omega} = \left[\begin{smallmatrix} \Omega_{uu} & \boldsymbol{\Omega}_{uv} \\ \boldsymbol{\Omega}_{vu} & \boldsymbol{\Omega}_{vv} \end{smallmatrix}\right]$. The one-sided long-run covariance matrix $\boldsymbol{\Delta} = \sum_{h=0}^{\infty} \mathbb{E}\left(\left[\begin{smallmatrix} u_t u_{t+h} & u_t \boldsymbol{v}_{t+h}' \\ \boldsymbol{v}_t u_{t+h} & \boldsymbol{v}_t \boldsymbol{v}_{t+h}' \end{smallmatrix}\right]\right) = \left[\begin{smallmatrix} \Delta_{uu} & \boldsymbol{\Delta}_{uv} \\ \boldsymbol{\Delta}_{vu} & \boldsymbol{\Delta}_{vv} \end{smallmatrix}\right]$

is partitioned similarly. Subscripts are used to refer to specific elements. For example, $\boldsymbol{B}_{v_i}$ and $\boldsymbol{\Delta}_{v_i u}$ denote the i^{th} elements of $\boldsymbol{B}_v$ and $\boldsymbol{\Delta}_{vu}$, respectively.

A concise exposition of our results asks for additional notation. An enumeration of various definitions is presented below.

(1) Introduce scaling matrices: $\boldsymbol{D}_{d,T}(\boldsymbol{\theta}) = \text{diag}[T^{\theta_1}, T^{\theta_2}, \ldots, T^{\theta_d}]$ for the time trends and their coefficients, and $\boldsymbol{D}_{s,T} = \text{diag}[\boldsymbol{D}_{(1),T}, \ldots, \boldsymbol{D}_{(m),T}]$ for the integer powers of $I(1)$ regressors, where $\boldsymbol{D}_{(i),T} = \text{diag}[T^{1/2}, T, \ldots, T^{p_i/2}]$. Moreover, we define two $(2d+p)\times(2d+p)$ nonrandom block matrices $\boldsymbol{L}_{\boldsymbol{\tau}_0,T}$ and $\boldsymbol{D}_{\boldsymbol{\theta}_0,T}$ such that

$$\boldsymbol{L}_{\boldsymbol{\tau}_0,T} = \begin{bmatrix} \boldsymbol{I}_d & -\text{diag}[\boldsymbol{\tau}_0]\ln T & \\ & \boldsymbol{I}_d & \\ & & \boldsymbol{I}_p \end{bmatrix}, \quad \boldsymbol{D}_{\boldsymbol{\theta}_0,T} = \sqrt{T}\begin{bmatrix} \boldsymbol{D}_{d,T}(\boldsymbol{\theta}_0) & & \\ & \boldsymbol{D}_{d,T}(\boldsymbol{\theta}_0) & \\ & & \boldsymbol{D}_{s,T} \end{bmatrix},$$

and

$$\boldsymbol{G}_{\gamma_0,T} = \boldsymbol{D}_{\boldsymbol{\theta}_0,T}\boldsymbol{L}'^{-1}_{\boldsymbol{\tau}_0,T} = \sqrt{T}\begin{bmatrix} \boldsymbol{D}_{d,T}(\boldsymbol{\theta}_0) & & \\ \boldsymbol{D}_{d,T}(\boldsymbol{\theta}_0)\,\text{diag}[\boldsymbol{\tau}_0]\ln T & \boldsymbol{D}_{d,T}(\boldsymbol{\theta}_0) & \\ \mathbf{O}_{p\times d} & \mathbf{O}_{p\times d} & \boldsymbol{D}_{s,T} \end{bmatrix}.$$

(2) Define vectors $\boldsymbol{d}(r;\boldsymbol{\theta}_0) = [r^{\theta_{10}}, r^{\theta_{20}}, \ldots, r^{\theta_{d0}}]'$, $\boldsymbol{B}_{(i)}(r) = [\boldsymbol{B}_{v_i}(r), \boldsymbol{B}^2_{v_i}(r), \ldots, \boldsymbol{B}^{p_i}_{v_i}(r)]'$ and their stacked random vector process $\boldsymbol{j}(r;\gamma_0) = [\,(\boldsymbol{\tau}_0 \odot \boldsymbol{d}(r;\boldsymbol{\theta}_0))'\ln(r), \boldsymbol{d}(r;\boldsymbol{\theta}_0)', \boldsymbol{B}'_{(1)}(r), \ldots, \boldsymbol{B}'_{(m)}(r)]'$.

(3) For the second-order bias terms, we define $\boldsymbol{b}_i = \left[1, 2\int_0^1 \boldsymbol{B}_{v_i}(r)dr, \ldots, p_i\int_0^1 \boldsymbol{B}^{p_i-1}_{v_i}(r)dr\right]'$ and $\boldsymbol{\mathcal{B}}_{vu} = [0'_{d\times1}, 0'_{d\times1}, \boldsymbol{b}'_1\boldsymbol{\Delta}_{v_1 u}, \ldots, \boldsymbol{b}'_m\boldsymbol{\Delta}_{v_m u}]'$.

Theorem 1

Under Assumptions 1-2, we have

$$\boldsymbol{G}_{\gamma_0,T}(\widehat{\gamma}_T - \gamma_0) \Rightarrow \left(\int \boldsymbol{j}(r;\gamma_0)\boldsymbol{j}(r;\gamma_0)'dr\right)^{-1}\left(\int \boldsymbol{j}(r;\gamma_0)dB_u(r) + \boldsymbol{\mathcal{B}}_{vu}\right), \quad \text{as } T \to \infty.$$

The proof of Theorem 1 is closely related to the work by Chan and Wang (2015). These authors provide the asymptotic distribution of NLS estimators with nonstationary time series under a set of general conditions (see their theorem 3.1). We verify that these conditions are also fulfilled when the scaling matrix $\boldsymbol{G}_{\gamma_0,T}$ is non-diagonal and depending on the true parameter vector γ_0. The results in Chan and Wang (2015) and Wang et al. (2018) suggest that Assumption 2 can be replaced by a long memory specification for $\Delta\boldsymbol{x}_t$. However, long memory parameters will enter the limiting distribution and inference will be complicated further.

We now illustrate Theorem 1 with two examples. These examples highlight two mathematical features that will complicate parameter inference.

Example 1

We consider the model $y_t = \tau t^\theta + u_t$ where the innovations satisfy Assumption 2. The limiting distribution of the parameter estimators depends solely on the mean square Riemann-Stieltjes *integrals* $\int \tau_0 r^{\theta_0}\ln(r)dB_u$ *and* $\int r^{\theta_0}dB_u$ *and is therefore normally distributed (e.g., section 2.3 in Tanaka (2017)):*

$$\begin{bmatrix} T^{\theta_0+\frac{1}{2}} & 0 \\ T^{\theta_0+\frac{1}{2}}\tau_0\ln(T) & T^{\theta_0+\frac{1}{2}} \end{bmatrix}\begin{bmatrix} \widehat{\theta}_T-\theta_0 \\ \widehat{\tau}_T-\tau_0 \end{bmatrix} \longrightarrow_d \mathrm{N}\left(\mathbf{0}, \Omega_{uu}(2\theta_0+1)^3\begin{bmatrix} 2\tau_0^2 & -\tau_0(2\theta_0+1) \\ -\tau_0(2\theta_0+1) & (2\theta_0+1)^2 \end{bmatrix}^{-1}\right). \quad (3.2)$$

The scaling matrix in the LHS of (3.2), $\begin{bmatrix} T^{\theta_0+\frac{1}{2}} & 0 \\ T^{\theta_0+\frac{1}{2}}\tau_0\ln(T) & T^{\theta_0+\frac{1}{2}} \end{bmatrix}$, *depends on* θ_0 *and is non-diagonal.*

Example 2

If $y_t = \tau t^{\theta} + \phi x_t + u_t$, *then the limiting distribution of the NLS estimator is:*

$$\begin{bmatrix} T^{\theta_0+\frac{1}{2}} & & \\ T^{\theta_0+\frac{1}{2}}\tau_0\ln(T) & T^{\theta_0+\frac{1}{2}} & \\ & & T \end{bmatrix}\begin{bmatrix} \widehat{\theta}_T-\theta_0 \\ \widehat{\tau}_T-\tau_0 \\ \widehat{\phi}_T-\phi_0 \end{bmatrix} \Rightarrow \begin{bmatrix} \int(\tau_0 r^{\theta_0}\ln(r))^2dr & \int\tau_0 r^{2\theta_0}\ln(r)dr & \int\tau_0 r^{\theta_0}\ln(r)B_vdr \\ \int\tau_0 r^{2\theta_0}\ln(r)dr & \int r^{2\theta_0}dr & \int r^{\theta_0}B_vdr \\ \int\tau_0 r^{\theta_0}\ln(r)B_vdr & \int r^{\theta_0}B_vdr & \int B_v^2dr \end{bmatrix}^{-1} \times$$

$$\left(\begin{bmatrix} \int\tau_0 r^{\theta_0}\ln(r)dB_u \\ \int r^{\theta_0}dB_u \\ \int B_vdB_u \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ \Delta_{vu} \end{bmatrix}\right).$$

This limiting distribution exhibits second order bias when $\Delta_{vu} \neq 0$, *or when* B_u *and* B_v *are correlated.*

Two features of the limiting distribution of $\boldsymbol{G}_{\gamma_0,T}(\widehat{\gamma}_T - \gamma_0)$ deserve further comments. First, as emphasised in Example 1, the scaling matrix $\boldsymbol{G}_{\gamma_0,T}$ features two uncommon properties: (1) this matrix depends on the true parameter vectors $\boldsymbol{\tau}_0$ and $\boldsymbol{\theta}_0$, and (2) $\boldsymbol{G}_{\gamma_0,T}$ is not diagonal. These peculiarities are caused by the nonlinearity and nonstationarity of the model. More specifically, these features can be traced back to the presence of functions like $f(t;\tau,\theta) = \tau t^{\theta}$. Limiting distributions with a similar mathematical structure can be found in the structural breaks literature, cf. model setting II.b of Perron and Zhu (2005) and its detailed analysis in Beutner et al. (2020).

Second, the nonstationary regressor x_{it} enters the model (2.1) through a polynomial transformation of the form $g(x,\phi_i) = \phi_{i1}x + \phi_{i2}x^2 + \ldots + \phi_{ip_i}x^{p_i}$ $(i = 1,2,\ldots,m)$. In the terminology of Park and Phillips (2001) this part of the regression function is a linear combination of H_0-regular functions. It is well-documented in the literature, e.g. Chang et al. (2001) and Chan and Wang (2015), that this leads to second order bias terms and hence nonstandard inference (except in the special case of strictly exogenous nonstationary regressors).

Remark 3

Asymptotic results with a diagonal scaling matrix can be obtained at the expense of a singular joint limiting distribution. For example, reconsider Example 1 and note that $\begin{bmatrix} T^{\theta_0+\frac{1}{2}} & 0 \\ 0 & T^{\theta_0+\frac{1}{2}}/\ln(T) \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ -\tau_0 & 1/\ln(T) \end{bmatrix}\begin{bmatrix} T^{\theta_0+\frac{1}{2}} & 0 \\ T^{\theta_0+\frac{1}{2}}\tau_0\ln(T) & T^{\theta_0+\frac{1}{2}} \end{bmatrix}$. *Since* $\lim_{T\to\infty}\begin{bmatrix} 1 & 0 \\ -\tau_0 & 1/\ln(T) \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ -\tau_0 & 0 \end{bmatrix}$, *the continuous mapping theorem implies*

$$\begin{bmatrix} T^{\theta_0+\frac{1}{2}} & 0 \\ 0 & T^{\theta_0+\frac{1}{2}}/\ln(T) \end{bmatrix}\begin{bmatrix} \widehat{\theta}_T-\theta_0 \\ \widehat{\tau}_T-\tau_0 \end{bmatrix} \longrightarrow_d \begin{bmatrix} 1/\tau_0 \\ -1 \end{bmatrix} \times \mathrm{N}\left(\mathbf{0}, \Omega_{uu}(2\theta_0+1)^3\right),$$

and we recover the limiting distribution reported in theorem 6.3 of Phillips (2007).

3.1 General Considerations

Let us assume for the moment that $\boldsymbol{\theta}_0$ is known. The resulting model, that is $y_t = \boldsymbol{d}_t(\boldsymbol{\theta}_0)'\boldsymbol{\tau} + \boldsymbol{s}_t'\boldsymbol{\phi} + u_t$, is now linear in the unknown parameters $[\boldsymbol{\tau}', \boldsymbol{\phi}']'$. The OLS estimator is $\left[\begin{smallmatrix} \widetilde{\boldsymbol{\tau}}_T(\boldsymbol{\theta}_0) \\ \widetilde{\boldsymbol{\phi}}_T(\boldsymbol{\theta}_0) \end{smallmatrix}\right] = \left(\sum_{t=1}^T \boldsymbol{z}_t(\boldsymbol{\theta}_0)\boldsymbol{z}_t(\boldsymbol{\theta}_0)'\right)^{-1} \left(\sum_{t=1}^T \boldsymbol{z}_t(\boldsymbol{\theta}_0)y_t\right)$ and parameter inference is relatively straightforward. For example, we can apply the Fully Modified (FM) corrections as in Wagner and Hong (2016) to obtain a zero-mean Gaussian mixture limiting distribution that allows for standard inference.[6]

Given that $\boldsymbol{\theta}_0$ is unknown in practice, it seems natural to first compute $\widehat{\boldsymbol{\theta}}_T$ by minimisation of $\widetilde{Q}_T(\boldsymbol{\theta})$, and to subsequently compute

$$\begin{bmatrix} \widehat{\boldsymbol{\tau}}_T \\ \widehat{\boldsymbol{\phi}}_T \end{bmatrix} = \left(\sum_{t=1}^T \boldsymbol{z}_t(\widehat{\boldsymbol{\theta}}_T)\boldsymbol{z}_t(\widehat{\boldsymbol{\theta}}_T)'\right)^{-1} \sum_{t=1}^T \boldsymbol{z}_t(\widehat{\boldsymbol{\theta}}_T)y_t. \tag{3.3}$$

The latter estimator is linear in $y_1, y_2, \ldots, y_T$ and fully modified adjustments seem possible. However, there are two issues. First, this estimator does not allow us to conduct inference on $\boldsymbol{\theta}$. Second, it is not completely clear how the estimation error in $\widehat{\boldsymbol{\theta}}_T$ influences the limiting results.[7]

There is also some good news. Namely, if the estimator in (3.3) is used to calculate residuals, then these residuals can be used to construct consistent kernel estimators for the long-run variance (LRV) matrices $\boldsymbol{\Delta}$ and $\boldsymbol{\Omega}$. With $\boldsymbol{V}_t(\gamma) = [y_t - \boldsymbol{d}_t(\boldsymbol{\theta})'\boldsymbol{\tau} - \boldsymbol{s}_t'\boldsymbol{\phi}, \Delta\boldsymbol{x}_t']'$, these estimators are defined as

$$\widehat{\boldsymbol{\Delta}}_T = \frac{1}{T}\sum_{t=1}^T\sum_{s=1}^t k\left(\frac{|t-s|}{b_T}\right)\boldsymbol{V}_t(\widehat{\gamma}_T)\boldsymbol{V}_t(\widehat{\gamma}_T)', \qquad \widehat{\boldsymbol{\Omega}}_T = \frac{1}{T}\sum_{t=1}^T\sum_{s=1}^T k\left(\frac{|t-s|}{b_T}\right)\boldsymbol{V}_t(\widehat{\gamma}_T)\boldsymbol{V}_t(\widehat{\gamma}_T)', \tag{3.4}$$

for some kernel function $k(\cdot)$ and bandwidth parameter b_T. The first element in $\boldsymbol{V}_t(\widehat{\gamma}_T)$ is indeed the residual $\hat{u}_t = y_t - \boldsymbol{d}_t(\widehat{\boldsymbol{\theta}}_T)'\widehat{\boldsymbol{\tau}}_T - \boldsymbol{s}_t'\widehat{\boldsymbol{\phi}}_T$. The remaining elements are $\Delta\boldsymbol{x}_t = \boldsymbol{v}_t$.

Assumption 3

(a) $k(0) = 1$, $k(\cdot)$ *is continuous at zero, and* $\sup_{x\geq 0}|k(x)| < \infty$.

(b) $\int_0^\infty \bar{k}(x)dx < \infty$, *where* $\bar{k}(x) = \sup_{y\geq x}|k(y)|$.

(c) The bandwidth parameters $\{b_T : T \geq 1\}$ *satisfies* $\{b_T\} \subseteq (0,\infty)$ *as well as* $\lim_{T\to\infty}\left(b_T^{-1} + T^{-1/2}b_T \ln T\right) = 0$.

The conditions on the kernel function $k(\cdot)$, Assumptions 3(a)-(b), are identical to those in Jansson (2002). Jansson (2002) remarks that these assumptions "*would appear to be satisfied by any kernel in actual use*". Commonly used kernel functions such as the Bartlett, Parzen, and Quadratic Spectral kernels indeed satisfy all these assumptions. Assumption 3(c) differs from the usual requirement, $\lim_{T\to\infty}\left(b_T^{-1} + T^{-1/2}b_T\right) = 0$, by a factor $\ln T$. The difference is due to the estimation error in $\widehat{\boldsymbol{\theta}}_T$. This error causes the residuals $\{\hat{u}_t\}$ to be less close to the innovations $\{u_t\}$ and we balance this by including autocovariance matrices of higher lags at a lower rate.

[6]The deterministic component in the CPR model by Wagner and Hong (2016) is a linear combination of the elements in the vector $[1, t, t^2, \ldots, t^q]'$. The FM corrections are thus immediate if $\boldsymbol{\theta}_0$ is known and takes values in the natural numbers. Based on Lemma 2 in the Appendix, it is also relatively straightforward to derive such corrections when $\boldsymbol{\theta}_0$ is known but not necessarily elements of the natural numbers.

[7]We have investigated the asymptotic behaviour of a Fully Modified version of the estimator in (3.3). Our efforts in bounding the estimation error of $\widehat{\boldsymbol{\theta}}_T$ lead to a term in the covariance asymptotics that is $O_p(\ln T)$ instead of $o_p(1)$. This (as well as our Monte Carlo simulations) suggests that this Fully Modified estimator is not asymptotically valid.

Theorem 2
Under Assumptions 1-3, we have $\widehat{\boldsymbol{\Delta}}_T \longrightarrow_p \boldsymbol{\Delta}$ *and* $\widehat{\boldsymbol{\Omega}}_T \longrightarrow_p \boldsymbol{\Omega}$.

3.2 Simulated inference

The limiting distribution in Theorem 1 is nonpivotal and thus unsuited for inference. We will use a simulated approach to account for the nuisance parameters, i.e. $\boldsymbol{\tau}_0$, $\boldsymbol{\theta}_0$ and the parameters describing the covariance structure of $\boldsymbol{B}(r)$. The main idea is to replace nuisance parameters by consistent estimates and to rely on a Monte Carlo (MC) simulation to approximate the limiting distribution. The empirical quantiles of these MC draws allow us to test hypothesis and/or conduct inference. Clearly, this kind of approach will provide exact inference when the limiting distribution is invariant with respect to the nuisance parameters (e.g. Dufour and Khalaf (2002) and Dufour (2006)). In the absence of such invariance, results as those in Wang et al. (2018) and Bergamelli et al. (2019) show that the simulation approach can retain an asymptotic justification. The following algorithm is an adaptation of Wang et al.'s (2018) simulated estimation. Among others, the current setting has to control for more nuisance parameters because of the flexible trend specification.

Step 1: Estimate $\widehat{\gamma}_T$ and use the residuals $\{\hat{u}_t\}$ to compute the estimators $\widehat{\boldsymbol{\Delta}}_T$ and $\widehat{\boldsymbol{\Omega}}_T$ from (3.4)

Step 2: Repeat for $j = 1, \ldots, J$,

(a) Draw an $(m+1)$-dimensional sequence $\{\boldsymbol{e}_n\}_{n=1}^N$ i.i.d. from $\mathrm{N}(\boldsymbol{0}, \boldsymbol{I}_{m+1})$.

(b) Compute $\left[\begin{smallmatrix} \widehat{\mu}_n \\ \widehat{\boldsymbol{v}}_n \end{smallmatrix}\right] = \widehat{\boldsymbol{\Omega}}_T^{1/2} \boldsymbol{e}_n$ and construct the partial sum process $\{\widehat{\boldsymbol{\chi}}_n = [\widehat{\chi}_{1n}, \ldots, \widehat{\chi}_{mn}]'\}_{n=1}^N$ according to $\widehat{\boldsymbol{\chi}}_n = \widehat{\boldsymbol{\chi}}_{n-1} + \widehat{\boldsymbol{v}}_n$ and $\widehat{\boldsymbol{\chi}}_0 = \boldsymbol{0}$.

(c) Set $\widehat{\boldsymbol{w}}_n = [n, \widehat{\boldsymbol{\chi}}_n']'$, and construct a simulated draw as:

$$\widehat{\boldsymbol{\mathcal{J}}}_N^{(j)}\left(\widehat{\gamma}_T, \widehat{\boldsymbol{\Omega}}_T, \widehat{\boldsymbol{\Delta}}_{vu}^{-}\right) = \left\{\boldsymbol{G}_{\widehat{\gamma},N}'^{-1}\left[\sum_{n=1}^N \dot{f}(\widehat{\boldsymbol{w}}_n, \widehat{\gamma}_T)\dot{f}(\widehat{\boldsymbol{w}}_n, \widehat{\gamma}_T)'\right]\boldsymbol{G}_{\widehat{\gamma},N}^{-1}\right\}^{-1}\left\{\boldsymbol{G}_{\widehat{\gamma},N}'^{-1}\left[\sum_{n=1}^N \dot{f}(\widehat{\boldsymbol{w}}_n, \widehat{\gamma}_T)\widehat{\mu}_n\right] + \widehat{\boldsymbol{\mathcal{B}}}_{vu}^{-}\right\},$$

where $\widehat{\boldsymbol{\Delta}}_{vu}^{-}$ is a consistent estimator of the subblock of $\boldsymbol{\Delta}^{-} = \left[\begin{smallmatrix} \boldsymbol{\Delta}_{uu}^{-} & \boldsymbol{\Delta}_{uv}^{-} \\ \boldsymbol{\Delta}_{vu}^{-} & \boldsymbol{\Delta}_{vv}^{-} \end{smallmatrix}\right] = \boldsymbol{\Sigma} - \boldsymbol{\Delta}'$, and $\widehat{\boldsymbol{\mathcal{B}}}_{vu}^{-} = \left[\boldsymbol{0}_{d\times 1}', \boldsymbol{0}_{d\times 1}', \widehat{\boldsymbol{b}}_1'\widehat{\boldsymbol{\Delta}}_{v_1 u}^{-}, \ldots, \widehat{\boldsymbol{b}}_m'\widehat{\boldsymbol{\Delta}}_{v_m u}^{-}\right]'$ with $\widehat{\boldsymbol{b}}_i = \left[1, 2\frac{1}{N}\sum_{n=1}^N\left(\frac{\widehat{\chi}_{in}}{\sqrt{N}}\right), \ldots, p_i\frac{1}{N}\sum_{n=1}^N\left(\frac{\widehat{\chi}_{in}}{\sqrt{N}}\right)^{p_i-1}\right]'$.

Step 3: Use the empirical quantiles of elements of $\left\{\widehat{\boldsymbol{\mathcal{J}}}_N^{(1)}, \ldots, \widehat{\boldsymbol{\mathcal{J}}}_N^{(J)}\right\}$ to conduct inference.

Step 3 has been kept general for notational convenience. A more concrete example is as follows. To construct a two-sided equal-tailed confidence interval for θ_{01}, we calculate the $\frac{\alpha}{2}$ and $1 - \frac{\alpha}{2}$ empirical quantile of the first elements of $\left\{\widehat{\boldsymbol{\mathcal{J}}}_N^{(1)}, \ldots, \widehat{\boldsymbol{\mathcal{J}}}_N^{(J)}\right\}$, say $c_{\alpha/2}$ and $c_{1-\alpha/2}$ respectively. The implied confidence interval is $\left[\hat{\theta}_1 - c_{1-\alpha/2}T^{-(\hat{\theta}_1+\frac{1}{2})}, \hat{\theta}_1 - c_{\alpha/2}T^{-(\hat{\theta}_1+\frac{1}{2})}\right]$.

Theorem 3
Suppose Assumptions 1-3 hold and let $N = [\kappa T^{\alpha}]$ *for some* $\kappa > 0$ *and* $0 < \alpha \leq \min\{1, 1 + 2\tilde{\theta}\}$ *with* $\tilde{\theta} \in (-\frac{1}{2}, \theta_L)$. *Then, we have*

$$\left\{\boldsymbol{G}_{\widehat{\gamma},N}'^{-1}\left[\sum_{n=1}^N \dot{f}(\widehat{\boldsymbol{w}}_n, \widehat{\gamma}_T)\dot{f}(\widehat{\boldsymbol{w}}_n, \widehat{\gamma}_T)'\right]\boldsymbol{G}_{\widehat{\gamma},N}^{-1}\right\}^{-1}\left\{\boldsymbol{G}_{\widehat{\gamma},N}'^{-1}\left[\sum_{n=1}^N \dot{f}(\widehat{\boldsymbol{w}}_n, \widehat{\gamma}_T)\widehat{\mu}_n\right] + \widehat{\boldsymbol{\mathcal{B}}}_{vu}^{-}\right\}$$
$$\longrightarrow_{d^*} \left(\int \boldsymbol{j}(r;\gamma_0)\boldsymbol{j}(r;\gamma_0)'dr\right)^{-1}\left(\int \boldsymbol{j}(r;\gamma_0)dB_u(r) + \boldsymbol{\mathcal{B}}_{vu}\right),$$

in probability.

Theorem 3 establishes the asymptotic validity of the simulation approach. That is, for a large enough J, the empirical quantiles of the simulated distribution will coincide with the asymptotic distribution. According to Theorem 3, the length of the simulated time series, N, should grow more slowly as θ_L approaches $-\frac{1}{2}$. The actual choice of θ_L should satisfy $\theta_L < \theta_{0i}$ (for all $1 \leq i \leq d$).

3.3 KPSS-type test for the null of cointegration

The correct specification of the nonlinear cointegrating relation will result in a stationary error process $\{u_t\}_{t\in\mathbb{Z}}$. Stationarity tests can thus be used to detect spurious relationships and/or the omission of relevant terms from the cointegrating regression. We consider a KPSS-type test statistic for the null hypothesis of stationarity. The test statistic reads

$$K_T^+ = \widehat{\Omega}_{u.v}^{-1} \frac{1}{T} \sum_{t=1}^{T} \left(\frac{1}{\sqrt{T}} \sum_{i=1}^{T} \hat{u}_t^+ \right)^2, \tag{3.5}$$

where $\widehat{\Omega}_{u.v}$ is a consistent estimator of $\Omega_{u.v} = \Omega_{uu} - \boldsymbol{\Omega}_{uv}\boldsymbol{\Omega}_{vv}^{-1}\boldsymbol{\Omega}_{vu}$, $\hat{u}_t^+ = y_t^+ - \boldsymbol{d}_t(\widehat{\boldsymbol{\theta}}_T)' \, \widehat{\boldsymbol{\tau}}_T - \boldsymbol{s}_t' \, \widehat{\boldsymbol{\phi}}_T$, and $y_t^+ = y_t - \widehat{\boldsymbol{\Omega}}_{vu}\widehat{\boldsymbol{\Omega}}_{vv}^{-1}\Delta\boldsymbol{x}_t$. The statistic is (stochastically) bounded under the null hypothesis but diverges under the alternative. Several authors have reported model settings in which the asymptotic null distribution of K_T^+ is known, e.g. Kwiatkowski et al. (1992) and Wagner and Hong (2016).

The estimation of $\boldsymbol{\theta}$ contaminates the limiting distribution of (3.5) with nuisance parameters.[8] Choi and Saikkonen (2010), Wagner and Hong (2016), Jiang et al. (2019), and Lin and Reuvers (2020), have shown that subsampling can resolve this issue. We will follow their approach and use subsamples of size q_T to compute the test statistics.

Theorem 4
Under Assumptions 1-3 and if $\lim_{T\to\infty}\left(q_T^{-1} + (\ln T)\left(\frac{q_T}{T}\right)^{\theta_L+\frac{1}{2}}\right) = 0$, *then for any* $\ell \in \{1, \ldots, T-q_T+1\}$ *we have*

$$K_{q_T,\ell}^+ = \widehat{\Omega}_{u.v}^{-1} \frac{1}{q_T} \sum_{t=\ell}^{\ell+q_T-1} \left(\frac{1}{\sqrt{q_T}} \sum_{i=\ell}^{t} \hat{u}_i^+ \right)^2 \Rightarrow \int_0^1 [W(r)]^2 dr, \tag{3.6}$$

where $\widehat{\Omega}_{u.v}$ *is a consistent estimator of* $\Omega_{u.v} = \Omega_{uu} - \boldsymbol{\Omega}_{uv}\boldsymbol{\Omega}_{vv}^{-1}\boldsymbol{\Omega}_{vu}$ *(Theorem 2) and* $W(\cdot)$ *denotes a standard Brownian motion.*

Theorem 4 does not provide any guidance on the choices for the starting value ℓ and the subsample size q_T. First, for a given q_T, Choi and Saikkonen (2010) argue that the use of a single subsample (instead of all T observations) implies a significant loss of power. We follow their example and combine all $M = [T/q_T]$ subresidual series of length q_T using a Bonferroni procedure. That is, we create subresiduals series by selecting adjacent blocks of q_T residuals while alternating between the start and end of the sample. We calculate the KPSS-type test statistic for each subseries, say $K_1, \ldots, K_M$, and reject the null of stationarity at significance α whenever $\max\{K_1, \ldots, K_M\}$

[8]Proposition 5 in Wagner and Hong (2016) shows that the limiting distribution of K_T^+ is free of nuisance parameters if $\boldsymbol{\theta}_0$ is known and only a single integrated regressor occurs with integer powers greater than one. This result does not carry over to the current setting because of the estimation error in $\widehat{\boldsymbol{\theta}}_T$

exceeds $c_{\alpha/M}$ which is defined by $\mathbb{P}\left(\int [W(r)]^2 dr \geq c_{\alpha/M}\right) = \alpha/M$. Finally, we select the block size q_T using Romano and Wolf's (2001) minimum volatility rule. The approach is now completely data-driven.

4 Simulations

This section lists various Monte Carlo simulations showing that the asymptotic approximations from Section 3 provide useful guidance in finite samples. Further details on the implementation are as follows. We consider $T \in \{100, 200, 500\}$.[9] The long-run covariance matrices in (3.4) are computed using the Barlett kernel, $k(x) = 1 - |x|$ for $|x| \leq 1$ (and zero otherwise), and the bandwidth selection method described in Andrews (1991). Simulated limiting distributions are based on $J = 999$ replicates and we set $N = T$ (because $\theta_L = 0$ suffices in our settings). We test at 5% significance and report results based on 2.5×10^4 MC replications.

Foreshadowing the empirical application, we use the DGP

$$y_t = \tau_1 + \tau_2 t + \tau_3 t^\theta + \phi_1 x_t + \phi_2 x_t^2 + u_t, \tag{4.1}$$

where $x_t = \sum_{s=1}^t v_s$. The parameter values are $\theta = 2$, $\boldsymbol{\tau} = [\tau_1, \tau_2, \tau_3]' = [7, 0.05, -5 \times 10^{-4}]'$, and $\boldsymbol{\phi} = [5, \phi_2]'$. The disturbance vector $[u_t, v_t]'$ is generated from the VAR(1) specification[10]

$$\begin{bmatrix} u_t \\ v_t \end{bmatrix} = \boldsymbol{A} \begin{bmatrix} u_{t-1} \\ v_{t-1} \end{bmatrix} + \begin{bmatrix} \eta_t \\ \epsilon_t \end{bmatrix}, \qquad \begin{bmatrix} \eta_t \\ \epsilon_t \end{bmatrix} \overset{i.i.d.}{\sim} \mathrm{N}\left(\mathbf{0}, \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}\right). \tag{4.2}$$

In (4.2), we construct the autoregressive matrix $\boldsymbol{A}$ along the following two steps: (1) generate a (2×2) random matrix $\boldsymbol{U}$ from U[0, 1] to construct the orthogonal matrix $\boldsymbol{H} = \boldsymbol{U}(\boldsymbol{U}'\boldsymbol{U})^{-1/2}$, and (2) compute $\boldsymbol{A} = \boldsymbol{HLH}'$ using

(A) $\boldsymbol{L} = \mathrm{diag}[0, 0]$, (C) $\boldsymbol{L} = \mathrm{diag}[0.7, 0.5]$,

(B) $\boldsymbol{L} = \mathrm{diag}[0.5, 0.3]$, (D) $\boldsymbol{L} = \mathrm{diag}[0.9, 0.7]$.

Settings (A)–(D) gradually increase the serial correlation in the error processes. The parameter $\rho \in \{0.00, 0.25, 0.50\}$ governs the amount of endogeneity.

We conduct four simulation experiments. Our first simulation experiments relate to testing the null of *linear* cointegration, i.e. we test $H_0 : \phi_2 = 0$ versus its two-sided alternative $H_a : \phi_2 \neq 0$. The empirical size ($\phi_2 = 0$) is computed using four different estimators: (1) the NLS estimator with simulated critical values as described in Section 3.2 (SimNLS); the NLS estimator with simulated critical values and the true value for $\theta_0 = 2$ being provided (SimNLS(θ_0)); (3) an heuristic FMOLS estimator which uses $\widehat{\theta}$ but ignores its estimation error (FMOLS); and (4) the FMOLS estimator based on $\theta_0 = 2$ (FMOLS(θ_0)). The results are listed in Table 2. It is immediately clear that simulated critical values improve size control. We point out three other observations. First, we see that the empirical size is rather insensitive to changes in ρ, whereas the introduction of serial correlation makes the test (more) oversized. This behaviour is well-documented in simulation settings where θ is known and restricted to be a natural number, cf. Wagner and Hong (2016) or

[9]Figure 1 is an exception. This figure reports power curves and here we decided to use $T \in \{100, 200, 300\}$ instead. This reduces the computational burden and avoids large parameter ranges where empirical power is equal to one.

[10]We start the VAR recursions from $\left[\begin{smallmatrix} u_0 \\ v_0 \end{smallmatrix}\right] = \mathbf{0}$ and subsequently use a presample of 50 observations to reduce the influence of these initial values.

Lin and Reuvers (2020). Second, FMOLS shows the largest size distortions and this indeed calls its validity into question (also see footnote 7). Finally, we compare the two estimators that are informed about the quadratic deterministic trend: SimNLS(θ_0) and FMOLS(θ_0). The model is now linear in its parameters and NLS estimation is no longer necessary. That is, we find ourselves in the model specification previously analysed in detail by Wagner and Hong (2016). The comparison of SimNLS(θ_0) and FMOLS(θ_0) indicates that simulated inference is also advantageous in this setting.

The subsequent simulations are about testing power. We simulate power curves by varying ϕ_2 over the interval $(0, 0.15]$ (Figure 1). Since the outcomes are rather insensitive to changes in the endogeneity parameter ρ, we keep it fixed at $\rho = 0.50$. Throughout settings (A)–(D), empirical power increases monotonically with ϕ_2. There are also power gains from increasing the sample size. The latter fact is slightly distorted in setting (D) because the test is oversized when serial correlation is high. Overall, the behaviour of these power curves is as expected.

Table 3 reports the empirical coverage and average confidence interval (CI) length of 95% confidence intervals for θ. The coverage is always below the desired nominal level of 95%. Coverage can drop as low as 54% when the sample size is small ($T = 100$) and the high serial correlation scenario (D) is used. This lack of coverage is caused by the imprecise estimation of the long-run variance (LRV) matrices. If we provide the true values of the LRV matrices, see the rows labeled Coverage(Ω), then coverage is almost exactly 95% throughout all designs. As expected, the average width of the CIs decreases with sample size.

Finally, as a fourth set of Monte Carlo experiments, we look at the finite sample properties of the KPSS test (Table 4). Comparing KPSS and KPSS(θ_0), we see that knowledge of the true value of θ_0 is beneficial as it always brings the empirical size closer to 5%. This difference aside, our KPSS outcomes are comparable to the results reported in table 1 of Choi and Saikkonen (2010). That is, the Bonferroni correction leads to conservative tests for up to moderate levels of serial correlation. At high levels of serial correlation (Setting (D)) we approach unit root behaviour and the KPSS tests are oversized.

5 Empirical Application

We examine the evidence for an EKC for a collection of 18 countries over the period 1870-2014 ($T = 145$). Economic growth is measured by GDP and we use carbon dioxide (CO_2) emissions as a proxy for air pollution. The origin of these data is as follows. We used population and GDP data from the Maddison Project (see https://www.rug.nl/ggdc/historicaldevelopment/maddison/). Our carbon dioxide data are fossil-fuel CO_2 emissions as made available by the Carbon Dioxide Information Analysis Center (CDIAC, see https://cdiac.ess-dive.lbl.gov). Both GDP and CO_2 emissions are expressed per capita and subsequently log-transformed. In accordance with the notation of this paper, we will denote them by x_t and y_t, respectively. The same data set (or subsets thereof) has also been studied by Wagner (2015), Chan and Wang (2015), Wang et al. (2018), Wagner et al. (2019), and Lin and Reuvers (2020).[11] This conveniently allows us to compare results. All user choices (kernel specification, bandwidth selection, etc.) are kept the same as during the simulation study (see page 11).

Prior to the analysis of the econometric models we will discuss several features of the time

[11] The stationarity properties of the series have been extensively studied and commented on in these papers. We will not repeat this analysis but refer the interested reader to the Supplement where such results can be found.

series for Belgium (Figure 2).[12] An inverted U-shaped relationship between GDP and CO_2 (both in log per capita) is clearly visible in Figure 2(a) and results like these have triggered the research on the Environmental Kuznetz Curve. However, the heat map time indication also shows that time is almost monotonically increasing along the curve. Time effects - e.g. increasing environmental awareness, advances in sustainable technologies - can be valid alternative explanations for these nonlinearities and their omission can (falsely) exaggerate the influence of GDP. It is for this reason that we developed and analysed the Generalized Cointegrating Polynomial Regression (GCPR).

More evidence for the importance of time effects is available in Figure 2(b). This figure depicts the same per capita series after detrending.[13] The inverted U-shape is now (visually) less pronounced or even absent.

Finally, we consider two competing possibilities to extend the traditional linear cointegration specification: $y_t = \tau_1 + \tau_2 t + \phi_1 x_t + u_t$. This model does not account for any nonlinear behaviour over time and is therefore ill-suited to fit the data displayed in Figure 2(c). Cointegrating polynomial regressions use integer powers of x_t to describe the curvature over time. Following Hu et al. (2019) we can allow for an integrated regressor with a flexible power and estimate $y_t = \tau_1 + \tau_2 t + \phi_1 x_t + \phi_2 x_t^\theta + u_t$. The residual sum of squares (RSS) of the NLS estimator for this specification is shown in Figure 2(d). The absence of a minimum at $\theta = 2$ casts doubt on the commonly used quadratic specification in x_t. Moreover, the lack of any minimum might be interpreted as a sign that log per capita GDP is not the source of nonlinearity. Alternatively, we can opt for a flexible deterministic trend as in $y_t = \tau_1 + \tau_2 t + \tau_3 t^\theta + \phi x_t + u_t$. The RSS now exhibits a clear minimum, see Figure 2(e). Considerations like these motivate the use of GCPRs. We will argue in the next pages that this last model specification is well-suited to capture the important features of the pollution data.

We continue the empirical analysis with a comparison of three model specifications. All three models are of the form:

$$y_t = \tau_1 + \tau_2 t + \tau_3 t^\theta + \phi_1 x_t + \phi_2 x_t^2 + u_t. \tag{5.1}$$

Model (M1) is the specification above with $\tau_3 = 0$. This model specification (possibly with the additional constraint $\tau_2 = 0$) has been explored in various papers, e.g. Piaggio and Padilla (2012), Wagner (2015), and Wang et al. (2018). For this model specification (M1), an inverted-U relationship results when $\phi_1 > 0$ and $\phi_2 < 0$ and empirical evidence hereof is traditionally interpreted as the existence of the EKC. Moreover, if these coefficients have the correct signs, then the turning point - the level of economic growth at which environmental improvement starts - can be computed as $\exp(-\phi_1/2\phi_2)$. Model (M1) is restrictive in the sense that nonlinear time effects (clearly visible in Figure 2) can only be explaining using the term $\phi_2 x_t^2$.

The model specifications (M2) and (M3) include deterministic nonlinear time trends. For model (M2), we allow for $\tau_3 \neq 0$ but fix $\theta = 2$. The model in (5.1) without further restrictions is referred to as (M3). In the latter model, the NLS estimator for θ is computed by a grid search over the values $\Theta = [0.05, 0.95] \cup [1.05, 10]$ and simulated inference is used (Section 3.2). Table 5 depicts

[12]The data for Austria, Belgium, and Finland are mentioned in both Wagner (2015) and Wagner et al. (2019) to behave in line with the EKC. We discuss Belgium in the main text but the interested reader can find the same figures for Austria and Finland in supplement section H.3. The conclusions are the same.

[13]The outcomes of the Perron and Yabu (2009) test (see Supplement) indicate that log per capita GDP is likely to have a deterministic trend component. It is thus recommended to have a deterministic trend in the model for log per capita CO_2 emissions. We should thus look at the relationship between GDP and CO_2 emissions (in log per capita) after partialling out the effect of the linear trend.

how increasingly flexible nonlinear deterministic trends affect the parameter estimates for ϕ_1 and ϕ_2. Judging only by the signs of $\widehat{\phi}_1$ and $\widehat{\phi}_2$ (thus ignoring potential stationarity in the errors), the EKC exists for 17 out of 18, 9 out of 18, and 8 out of 18 countries for (M1), (M2), and (M3), respectively. Moreover, the significance of squared log per capita GDP (that is ϕ_2) reduces when nonlinear deterministic time trends are included. For model (M3), ϕ_2 is never significantly different from zero at a 10% level and evidence in favour of EKC becomes rather meagre. The results of the KPSS tests for these models can be found in Table 5 under "Stationarity tests". In general, the cointegrating relations seem well-specified except maybe for Belgium, Denmark, and UK.[14]

The insignificance of ϕ_2 in model (M3) suggests us to return to the model specification that was introduced earlier, namely

$$y_t = \tau_1 + \tau_2 t + \tau_3 t^{\theta} + \phi_1 x_t + u_t. \tag{M4}$$

Model (M4) specifies a linear cointegrating relation around a flexible time trend and does not incorporate nonlinear effects in log per capita GDP.[15] That is, the model specification does not allow for an EKC. As before, we check parameter estimates and test for stationarity of the error terms (the columns labeled "($M4$)" in Table 5 and Figures 2(f), 3 and 4). Some remarks concerning this final model specification are:

1. For Belgium, the fitted model reads

$$y_t = -0.049 + 0.0063t - 6.131 \times 10^{-6}\, t^{2.603} + 1.006 x_t + \hat{u}_t. \tag{5.2}$$

   The flexible power on the linear trend is estimated to be $\widehat{\theta} = 2.603$ resulting in nonlinear behaviour over time. Moreover, the negative coefficient in front of $t^{2.603}$ provides a contribution that is sloping down over time. If time effects are ignored, then a 1% increase in GDP will lead to an estimated 1.006% increase in fossil-fuel CO_2 emissions.

2. The outcomes of the KPSS test do not point towards a misspecified cointegrating relation (Table 5). The flexible deterministic trend is apparently sufficient to describe the nonlinear behaviour of log per capita CO_2 emissions over time, that is squared log per capita GDP is not needed in the model. Visual proof is found in Figures 2(a), 2(b) and 2(f) where the incorporation of increasingly flexible time effects is seen to remove any apparent nonlinear relationship between log per capita GDP and CO_2 emissions.

3. The estimates for θ and their confidence intervals are reported in Figure 3. Since the parameter space for θ is bounded below by $-\frac{1}{2}$, we have truncated its confidence interval at this value. This reflects the belief that values less than $-\frac{1}{2}$ are impossible (within our model setting). For Japan and Portugal, we find $\widehat{\theta} = 0.05$, i.e. a value at the boundary of Θ. For these two countries we suggest to omit the flexible trend altogether.

4. Figure 4 compares the fit of CPR model (M1) with the fit of the GCPR model (M4). The fit of both models is comparable for most of the time span. However, the fit of model (M4) is often better at the start and end of the sample, say 1870-1890.

[14] Deciding on the correct specification of the cointegrating relations for each of the 18 countries is implicitly a joint test. The interpretation of the individual outcomes therefore suffers from the multiple testing problem. A multivariate stationarity test is discussed in Lin and Reuvers (2020).

[15] Model specification (M4) has the additional advantage of being invariant to the possible presence of a drift component in log per capita GDP, also see footnote 13.

6 Summary and conclusion

In this paper we have extended the cointegrating polynomial regression (CPR) model of Wagner and Hong (2016) with power law deterministic trends. The unknown powers are estimated jointly with the parameters in the cointegrating relation. The limiting distribution is nonstandard because it involves a non-diagonal scaling matrix and the usual second order bias effects. We therefore suggest a simulation-based approach to conduct inference. The usual subsampling KPSS-type for stationarity of the innovations of the nonlinear cointegrating relation remains valid. Our results are supported by Monte Carlo simulation. The empirical application on the Environmental Kuznets Curve shows that a flexible trend can fully capture the nonlinearity in the data thereby making higher order powers of log per capita GDP redundant. Our resulting model is linear in log per capita GDP and suggests an alternative explanation in which time effects (e.g. technological progress, environmental awareness) cause the recent slowdown in pollution. Contrary to the opening quote in the introduction, our data does not suggest that air quality will benefit from economic growth. Finally, the narrative of this paper as well as the empirical application are centred around the Environmental Kuznets Curve. However, our model setting can also find application elsewhere.

References

Adams, R. A. and C. Essex (2016). *Calculus: A Complete Course*. Pearson Canada.

Andrews, D. W. and X. Cheng (2012). Estimation and inference with weak, semi-strong, and strong identification. *Econometrica 80*, 2153–2211.

Andrews, D. W. K. (1991). Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica 59*, 817–858.

Baek, Y. I., S. J. Cho, and P. C. B. Phillips (2015). Testing linearity using power transforms of regressors. *Journal of Econometrics 187*, 376–384.

Bergamelli, M., A. Bianchi, L. Khalaf, and G. Urga (2019). Combining p-values to test for multiple structural breaks in cointegrated regressions. *Journal of Econometrics 211*, 461–482.

Beutner, E., Y. Lin, and S. Smeekes (2020). GLS estimation and confidence sets for the date of a single break in models with trends. Working Paper.

Carson, R. T. (2009). The Environmental Kuznets Curve: seeking empirical regularity and theoretical structure. *Review of Environmental Economics and Policy 4*, 3–23.

Chan, N. and Q. Wang (2015). Nonlinear regressions with nonstationary time series. *Journal of Econometrics 185*, 182–195.

Chang, Y., J. Y. Park, and P. C. B. Phillips (2001). Nonlinear econometric models with cointegrated and deterministically trending regressors. *The Econometrics Journal 4*, 1–36.

Cho, J. S. and P. C. B. Phillips (2018). Sequentially testing polynomial model hypotheses using power transforms of regressors. *Journal of Applied Econometrics 33*, 141–159.

Choi, I. and P. Saikkonen (2010). Tests for nonlinear cointegration. *Econometric Theory 26*, 682–709.

Dasgupta, S., B. K. H. Wang, and D. Wheeler (2002). Confronting the Environmental Kuznets Curve. *Journal of Economic Perspectives 16*, 147–168.

Davies, R. B. (1977). Hypothesis testing when a nuisance parameter is present only under the alternative. *Biometrika 64*, 247–254.

Davies, R. B. (1987). Hypothesis testing when a nuisance parameter is present only under the alternative. *Biometrika 74*, 33–43.

Dufour, J.-M. (2006). Monte Carlo tests with nuisance parameters: A general approach to finite-sample inference and nonstandard asymptotics. *Journal of Econometrics 133*, 443–477.

Dufour, J.-M. and L. Khalaf (2002). Simulation based finite and large sample tests in multivariate regressions. *Journal of Econometrics 111*, 303–322.

Durlauf, S. N. and P. C. B. Phillips (1988). Trends versus random walks in time series analysis. *Econometrica 56*, 1333–1354.

Gillingham, K. and J. H. Stock (2018). The cost of reducing greenhouse gas emissions. *Journal of Economic Perspectives 32*, 53–72.

Grossman, G. M. and A. B. Krueger (1995). Economic growth and the environment. *The Quarterly Journal of Economics 110*, 353–377.

Hamilton, J. D. (1994). *Time Series Analysis*. Princeton University Press.

Hansen, B. E. (1992). Consistent covariance matrix estimation for dependent heterogeneous processes. *Econometrica 60*, 967–972.

Hong, S. H. and P. C. B. Phillips (2010). Testing linearity in cointegrating relations with an application to purchasing power parity. *Journal of Business & Economic Statistics 28*, 96–114.

Hu, Z., P. C. B. Phillips, and Q. Wang (2019). Nonlinear cointegrating power function regression with endogeneity. Working paper.

Jansson, M. (2002). Consistent covariance matrix estimation for linear processes. *Econometric Theory 18*, 1449–1459.

Jiang, B., Y. Lu, and J. Y. Park (2019). Testing for stationary at high frequency. *Journal of Econometrics 215*, 341–374.

Kwiatkowski, D., P. C. B. Phillips, P. Schmidt, and Y. Shin (1992). Testing the null hypothesis of stationarity against the alternative of a unit root: How sure are we that economic time series have a unit root? *Journal of Econometrics 54*, 159–178.

Lin, Y. and H. Reuvers (2020). Efficient estimation by fully modified GLS with an application to the Environmental Kuznets Curve. Working paper.

Nordhaus, W. D. (2014). The perils of the learning model for modeling endogenous technological change. *The Energy Journal 35*.

Panayotou, T. (1997). Demystifying the environmental kuznets curve: Turning a black box into a policy tool. *Environment and Development Economics 2*, 465–484.

Park, J. Y. and P. C. B. Phillips (2001). Nonlinear regressions with integrated time series. *Econometrica 69*, 117–161.

Perron, P. and T. Yabu (2009). Estimating deterministic trends with an integrated or stationary noise component. *Journal of Econometrics 151*, 56–69.

Perron, P. and X. Zhu (2005). Structural breaks with deterministic and stochastic trends. *Journal of Econometrics 129*, 65–119.

Phillips, P. C. B. (1986). Understanding spurious regressions in econometrics. *Journal of Econometrics 33*, 311–340.

Phillips, P. C. B. (1998). New tools for understanding spurious regressions. *Econometrica 66*, 1299–1325.

Phillips, P. C. B. (2007). Regression with slowly varying regressors and nonlinear trends. *Econometric Theory 23*, 557–614.

Phillips, P. C. B. and V. Solo (1992). Asymptotics for linear processes. *The Annals of Statistics 20*, 971–1001.

Piaggio, M. and E. Padilla (2012). Co2 emissions and economic activity: Heterogeneity across countries and non-stationary series. *Energy Policy 46*, 370–381.

Robinson, P. M. (2012). Inference on power law spatial trends. *Bernoulli 18*, 644–677.

Romano, J. P. and M. Wolf (2001). Subsampling intervals in autoregressive models with linear time trend. *Econometrica 69*, 1283–1314.

Soong, T. T. (1973). *Random Differential Equations in Science and Engineering*. Academic Press, Inc.

Stern, D. I. (2004). The rise and fall of the environmental kuznets curve. *World Development 32*, 1419–1439.

Stypka, O., M. Wagner, P. Grabarczyk, and R. Kawka (2017). The asymptotic validity of "standard" fully modified OLS estimation and inference in cointegrating polynomial regressions. Working Paper.

Tanaka, K. (2017). *Time Series Analysis: Nonstationary and Noninvertible Distribution Theory*. John Wiley & Sons.

Wagner, M. (2015). The Environmental Kuznets Curve, cointegration and nonlinearity. *Journal of Applied Econometrics 30*, 948–967.

Wagner, M., P. Grabarczyk, and S. H. Hong (2019). Fully modified OLS estimation and inference for seemingly unrelated cointegrating polynomial regressions and the Environmental Kuznets Curve for carbon dioxide emissions. *Journal of Econometrics 214*, 216–255.

Wagner, M. and S. H. Hong (2016). Cointegrating polynomial regressions: Fully modified OLS estimation and inference. *Econometric Theory 32*, 1289–1315.

Wang, Q., D. Wu, and K. Zhu (2018). Model checks for nonlinear cointegrating regression. *Journal of Econometrics 207*, 261–284.

Table 1: The empirical size (in %) of a t-test for $H_0 : \phi_2 \geq 0$ versus $H_a : \phi_2 < 0$ with model specification $y_t = \tau_1 + \tau_2 t + \phi_1 x_t + \phi_2 x_t^2 + u_t$. The variable x_t is generated as a random walk process.

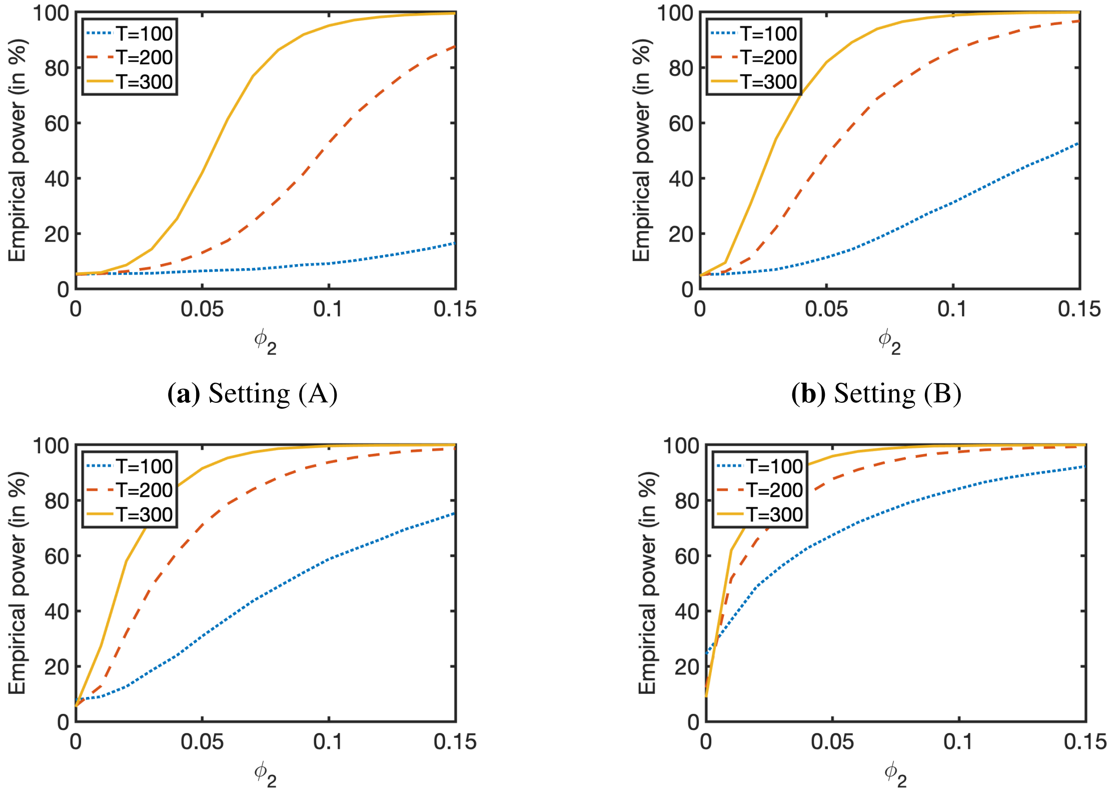
	DGP: $y_t = -\tau_0 t^2 + u_t$						
τ_0 (in 10^{-4})	0	1	2	3	4	5	6
$T = 100$	5.5	8.5	13.2	19.8	26.7	33.6	39.0
$T = 200$	5.5	37.4	57.9	65.2	68.6	70.6	72.5

Note: Further simulation details and a theoretical explanation of these Monte Carlo outcomes can be found in the Supplement. The mechanisms that cause this behaviour of the hypothesis test remind of the literature on spurious regressions and spurious detrending (for example Phillips (1986) and Durlauf and Phillips (1988), respectively).

Table 2: The empirical size (in %) of the coefficient test $H_0 : \phi_2 = 0$ versus $H_a : \phi_2 \neq 0$. The Monte Carlo results are based on: simulated inference with θ estimated by NLS (SimNLS), simulated inference with $\theta = 2$ (SimNLS(θ_0)), a Fully Modified estimator with estimated θ (FMOLS), and a Fully Modified estimator being informed about the true value of θ (FMOLS(θ_0)).

	(A)			(B)			(C)			(D)		
ρ	0.00	0.25	0.50	0.00	0.25	0.50	0.00	0.25	0.50	0.00	0.25	0.50
$T = 100$												
SimNLS	5.68	5.42	5.38	6.24	5.88	5.81	9.16	9.24	8.70	26.25	26.19	25.61
SimNLS(θ_0)	5.43	5.38	5.31	5.42	5.00	5.08	7.16	7.25	7.01	19.29	19.48	19.28
FMOLS	10.70	10.67	11.12	18.66	18.52	17.88	25.61	25.52	24.30	42.97	43.53	42.15
FMOLS(θ_0)	7.50	7.00	6.58	14.69	14.20	12.90	21.08	20.28	19.12	38.98	38.54	37.88
$T = 200$												
SimNLS	5.44	5.37	5.28	5.12	4.85	5.14	6.66	5.92	5.97	14.44	13.80	13.36
SimNLS(θ_0)	5.50	5.21	5.09	4.77	4.55	4.83	5.54	5.29	5.28	10.00	9.65	9.43
FMOLS	9.33	9.33	10.60	14.87	14.68	14.62	19.05	19.22	18.30	31.59	31.62	30.72
FMOLS(θ_0)	6.08	6.02	5.79	10.90	10.85	10.20	14.58	14.65	13.32	26.02	25.98	25.02
$T = 500$												
SimNLS	5.27	5.27	5.17	4.63	4.72	4.69	4.97	5.02	4.90	6.94	6.72	6.46
SimNLS(θ_0)	5.31	5.34	5.25	4.44	4.67	4.70	4.75	4.74	4.55	5.74	5.71	5.19
FMOLS	8.41	8.58	10.35	11.92	11.78	12.32	14.32	14.47	14.18	20.48	20.39	19.55
FMOLS(θ_0)	5.49	5.38	5.45	8.33	8.40	7.85	10.62	10.08	9.78	14.92	15.22	13.87

Note: The DGP is $y_t = \tau_1 + \tau_2 t + \tau_3 t^\theta + \phi_1 x_t + \phi_2 x_t^2 + u_t$ (numerical values are given on page 11) with $x_t = \sum_{s=1}^t v_s$. The innovations follow $\left[\begin{smallmatrix} u_t \\ v_t \end{smallmatrix}\right] = \boldsymbol{A}\left[\begin{smallmatrix} u_{t-1} \\ v_{t-1} \end{smallmatrix}\right] + \left[\begin{smallmatrix} \eta_t \\ \epsilon_t \end{smallmatrix}\right]$ with $\left[\begin{smallmatrix} \eta_t \\ \epsilon_t \end{smallmatrix}\right] \overset{i.i.d.}{\sim} \mathrm{N}\left(\boldsymbol{0}, \left[\begin{smallmatrix} 1 & \rho \\ \rho & 1 \end{smallmatrix}\right]\right)$. $\boldsymbol{A} = \boldsymbol{HLH}'$ with $\boldsymbol{H} = \boldsymbol{U}(\boldsymbol{U}'\boldsymbol{U})^{-1/2}$ and $\boldsymbol{U}$ being a (2×2) matrix of uniformly distributed random variables on $[0, 1]$. We consider four specifications for $\boldsymbol{L}$: (A) $\boldsymbol{L} = \mathrm{diag}[0, 0]$, (B) $\boldsymbol{L} = \mathrm{diag}[0.5, 0.3]$, (C) $\boldsymbol{L} = \mathrm{diag}[0.7, 0.5]$, and (D) $\boldsymbol{L} = \mathrm{diag}[0.9, 0.7]$.



(a) Setting (A) **(b)** Setting (B)

(c) Setting (C) **(d)** Setting (D)

Figure 1: The power curve for the test $H_0 : \phi_2 = 0$ versus its two-sided alternative $H_a : \phi_2 \neq 0$. The simulation DGP is $y_t = \tau_1 + \tau_2 t + \tau_3 t^\theta + \phi_1 x_t + \phi_2 x_t^2 + u_t$. Results are shown for settings (A)–(D) with $\rho = 0.50$. Simulation details can be found in Section 4.

Table 3: Simulation results on the confidence intervals for θ. We report the empirical coverage, the coverage(Ω) computed with the true LRVs, and average length of 95% confidence intervals. All computations use simulated inference, see Section 3.2.

	(A)			(B)			(C)			(D)		
ρ	0.00	0.25	0.50	0.00	0.25	0.50	0.00	0.25	0.50	0.00	0.25	0.50
$T = 100$												
Coverage	93.08	92.65	92.20	84.64	84.77	84.24	76.90	76.85	76.66	54.54	54.74	55.35
Coverage(Ω)	94.55	94.64	95.44	94.49	94.38	94.07	94.72	94.59	94.45	95.66	95.76	96.14
Length	0.66	0.70	0.81	0.87	0.89	0.92	1.14	1.14	1.16	1.76	1.79	1.76
$T = 200$												
Coverage	94.22	93.85	93.36	89.41	89.00	88.71	84.88	84.37	84.79	68.76	69.04	68.33
Coverage(Ω)	94.98	94.89	94.99	95.22	94.80	94.34	95.06	95.04	95.07	95.61	95.53	95.97
Length	0.12	0.13	0.15	0.17	0.17	0.18	0.24	0.24	0.24	0.39	0.39	0.39
$T = 500$												
Coverage	94.55	94.44	94.35	91.68	91.61	91.60	90.22	90.02	89.42	82.49	82.33	81.90
Coverage(Ω)	94.94	95.03	94.89	94.75	94.64	94.40	95.01	95.02	94.51	95.07	95.26	95.51
Length	0.01	0.01	0.02	0.02	0.02	0.02	0.03	0.03	0.03	0.06	0.06	0.06

Note: The DGP is $y_t = \tau_1 + \tau_2 t + \tau_3 t^\theta + \phi_1 x_t + \phi_2 x_t^2 + u_t$ (numerical values are given on page 11) with $x_t = \sum_{s=1}^t v_s$. The innovations follow $\left[\begin{smallmatrix} u_t \\ v_t \end{smallmatrix}\right] = \boldsymbol{A}\left[\begin{smallmatrix} u_{t-1} \\ v_{t-1} \end{smallmatrix}\right] + \left[\begin{smallmatrix} \eta_t \\ \epsilon_t \end{smallmatrix}\right]$ with $\left[\begin{smallmatrix} \eta_t \\ \epsilon_t \end{smallmatrix}\right] \overset{i.i.d.}{\sim} \mathrm{N}\left(\mathbf{0}, \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}\right)$. $\boldsymbol{A} = \boldsymbol{HLH}'$ with $\boldsymbol{H} = \boldsymbol{U}(\boldsymbol{U}'\boldsymbol{U})^{-1/2}$ and $\boldsymbol{U}$ being a (2×2) matrix of uniformly distributed random variables on $[0, 1]$. We consider four specifications for $\boldsymbol{L}$: (A) $\boldsymbol{L} = \mathrm{diag}[0, 0]$, (B) $\boldsymbol{L} = \mathrm{diag}[0.5, 0.3]$, (C) $\boldsymbol{L} = \mathrm{diag}[0.7, 0.5]$, and (D) $\boldsymbol{L} = \mathrm{diag}[0.9, 0.7]$.

Table 4: The empirical size (in %) of the subsampling Bonferroni KPSS tests. The row labeled 'KPSS' is computable in practice. We additionally report simulation outcomes on the same test when being informed about the true value of θ, see the row indicated by 'KPSS(θ_0)'.

	(A)			(B)			(C)			(D)		
ρ	0.00	0.25	0.50	0.00	0.25	0.50	0.00	0.25	0.50	0.00	0.25	0.50
$T = 100$												
KPSS	0.34	0.43	1.34	0.92	1.00	1.39	2.03	2.09	2.16	20.47	20.22	20.22
KPSS(θ_0)	0.69	0.78	1.90	1.32	1.46	1.89	2.30	2.23	2.23	16.48	16.51	16.42
$T = 200$												
KPSS	0.79	0.94	2.31	1.46	1.70	2.42	2.22	2.19	2.26	10.70	10.61	10.17
KPSS(θ_0)	1.19	1.24	2.69	1.82	2.08	2.70	2.43	2.38	2.50	8.92	8.86	8.24
$T = 500$												
KPSS	1.08	1.48	2.90	2.08	1.91	2.40	2.21	2.10	2.43	5.10	4.72	4.20
KPSS(θ_0)	1.36	1.76	3.13	2.27	2.24	2.56	2.32	2.27	2.54	4.35	3.98	3.68

Note: The DGP is $y_t = \tau_1 + \tau_2 t + \tau_3 t^\theta + \phi_1 x_t + \phi_2 x_t^2 + u_t$ with $\phi_2 = 0$ (and all other numerical values as on page 11) with $x_t = \sum_{s=1}^{t} v_s$. The innovations follow $\left[\begin{smallmatrix} u_t \\ v_t \end{smallmatrix}\right] = \boldsymbol{A}\left[\begin{smallmatrix} u_{t-1} \\ v_{t-1} \end{smallmatrix}\right] + \left[\begin{smallmatrix} \eta_t \\ \epsilon_t \end{smallmatrix}\right]$ with $\left[\begin{smallmatrix} \eta_t \\ \epsilon_t \end{smallmatrix}\right] \overset{i.i.d.}{\sim} \mathrm{N}\left(\mathbf{0}, \left[\begin{smallmatrix} 1 & \rho \\ \rho & 1 \end{smallmatrix}\right]\right)$. $\boldsymbol{A} = \boldsymbol{H}\boldsymbol{L}\boldsymbol{H}'$ with $\boldsymbol{H} = \boldsymbol{U}(\boldsymbol{U}'\boldsymbol{U})^{-1/2}$ and $\boldsymbol{U}$ being a (2×2) matrix of uniformly distributed random variables on $[0, 1]$. We consider four specifications for $\boldsymbol{L}$: (A) $\boldsymbol{L} = \mathrm{diag}[0, 0]$, (B) $\boldsymbol{L} = \mathrm{diag}[0.5, 0.3]$, (C) $\boldsymbol{L} = \mathrm{diag}[0.7, 0.5]$, and (D) $\boldsymbol{L} = \mathrm{diag}[0.9, 0.7]$.

Table 5: Parameter estimates and output of the KPSS-type of test for stationarity as computed for model specifications $(M1) - (M4)$. The column $\widehat{KPSS}$ and M_{opt} provide the numerical values of the KPSS tests and the number of chosen residual subblocks, respectively.

	Parameter estimates								Stationarity tests							
	$(M1)$		$(M2)$		$(M3)$		$(M4)$		$(M1)$		$(M2)$		$(M3)$		$(M4)$	
Country	$\widehat{\phi}_1$	$\widehat{\phi}_2$	$\widehat{\phi}_1$	$\widehat{\phi}_2$	$\widehat{\phi}_1$	$\widehat{\phi}_2$	$\widehat{\theta}$	$\widehat{\phi}_1$	$\widehat{KPSS}$	M_{opt}	$\widehat{KPSS}$	M_{opt}	$\widehat{KPSS}$	M_{opt}	$\widehat{KPSS}$	M_{opt}
Australia	2.75	−0.17	-23.92^{***}	1.40^{***}	-12.19^{***}	0.74	0.88	1.25^{***}	1.49	9	1.79	9	1.80	9	1.35	9
Austria	7.13^{***}	-0.30^{**}	1.25	0.03	3.75^{***}	−0.12	0.88	1.55^{***}	1.03	7	1.07	7	1.65	7	1.63	7
Belgium	11.45^{***}	-0.57^{***}	10.03^{***}	-0.49^{***}	10.29^{***}	−0.50	2.60	1.01^{***}	1.91	9	2.92^{*}	8	2.74^{*}	8	2.28	9
Canada	12.72^{***}	-0.64^{***}	14.80	−0.77	-3.46^{***}	0.25	0.56	1.14^{***}	2.83^{*}	7	2.60^{*}	7	1.26	9	1.28	9
Denmark	14.52^{***}	-0.65^{***}	−2.80	0.25	-5.75^{***}	0.39	2.03	1.68^{***}	3.14^{*}	9	3.30^{**}	9	2.98^{*}	9	1.58	8
Finland	16.86^{***}	-0.76^{***}	16.97^{***}	-0.77^{***}	22.58^{***}	−1.06	1.87	3.95^{***}	2.05	8	2.06	8	0.71	9	0.80	9
France	10.87^{***}	-0.55^{***}	3.14^{*}	−0.12	3.31^{***}	−0.13	2.09	1.00^{***}	1.72	9	0.69	8	0.56	8	2.49	9
Germany	6.24^{***}	-0.31^{***}	−1.82	0.13	-4.42^{***}	0.29	0.59	0.89^{***}	2.63^{*}	7	2.1	8	1.23	9	2.81^{*}	7
Italy	11.76^{***}	-0.55^{***}	7.31^{**}	−0.30	7.72^{***}	−0.29	0.82	2.41^{***}	4.18^{**}	7	3.79^{**}	8	0.79	7	0.78	7
Japan	9.86^{***}	-0.52^{***}	−4.27	0.29	1.16^{***}	−0.00	0.05	1.15^{***}	5.17^{***}	8	3.93^{**}	7	1.83	9	1.84	9
Netherlands	8.70^{***}	-0.41^{***}	1.49	−0.01	0.48^{**}	0.05	1.86	1.32^{***}	2.16	7	0.94	7	1.2	7	1.15	7
Norway	3.87	−0.16	-9.14^{**}	0.51^{**}	-1.10^{**}	0.16	0.46	2.05^{***}	2.53^{*}	7	1.06	7	0.74	9	1.44	8
Portugal	0.09	0.04	-5.86^{***}	0.42	-1.11^{**}	0.15	0.05	1.69^{***}	6.95^{***}	8	5.28^{**}	7	0.64	7	1.64	7
Spain	7.72^{***}	-0.37^{***}	1.98	−0.01	4.31^{***}	−0.16	1.55	1.52^{***}	2.78^{*}	7	2.03	8	2.4	8	2.42	8
Sweden	10.91^{***}	-0.44^{***}	-9.08^{*}	0.61^{**}	0.43	0.17	0.46	3.48^{***}	3.59^{**}	7	1.27	7	0.73	7	0.80	7
Switzerland	8.57^{***}	-0.29^{***}	-7.86^{**}	0.54^{***}	-13.86^{***}	0.83	2.98	2.63^{***}	0.80	7	0.96	7	0.77	7	0.75	7
UK	9.32^{***}	-0.47^{***}	5.91^{***}	-0.27^{***}	4.13^{***}	−0.18	3.04	0.80^{***}	2.76^{*}	9	4.25^{**}	9	4.30^{**}	9	3.98^{**}	9
USA	8.67^{***}	-0.44^{***}	0.93	−0.03	-5.62^{***}	0.35	0.92	0.95^{***}	1.64	8	1.85	8	2.25	8	1.97	8

Note: Asterisks denote rejection of the null hypothesis at the $^{***}1\%$, $^{**}5\%$, and $^{*}10\%$ significance level. Depending on the specific table entry, the null hypothesis refers to either a coefficient being zero or (nonlinear) cointegration.

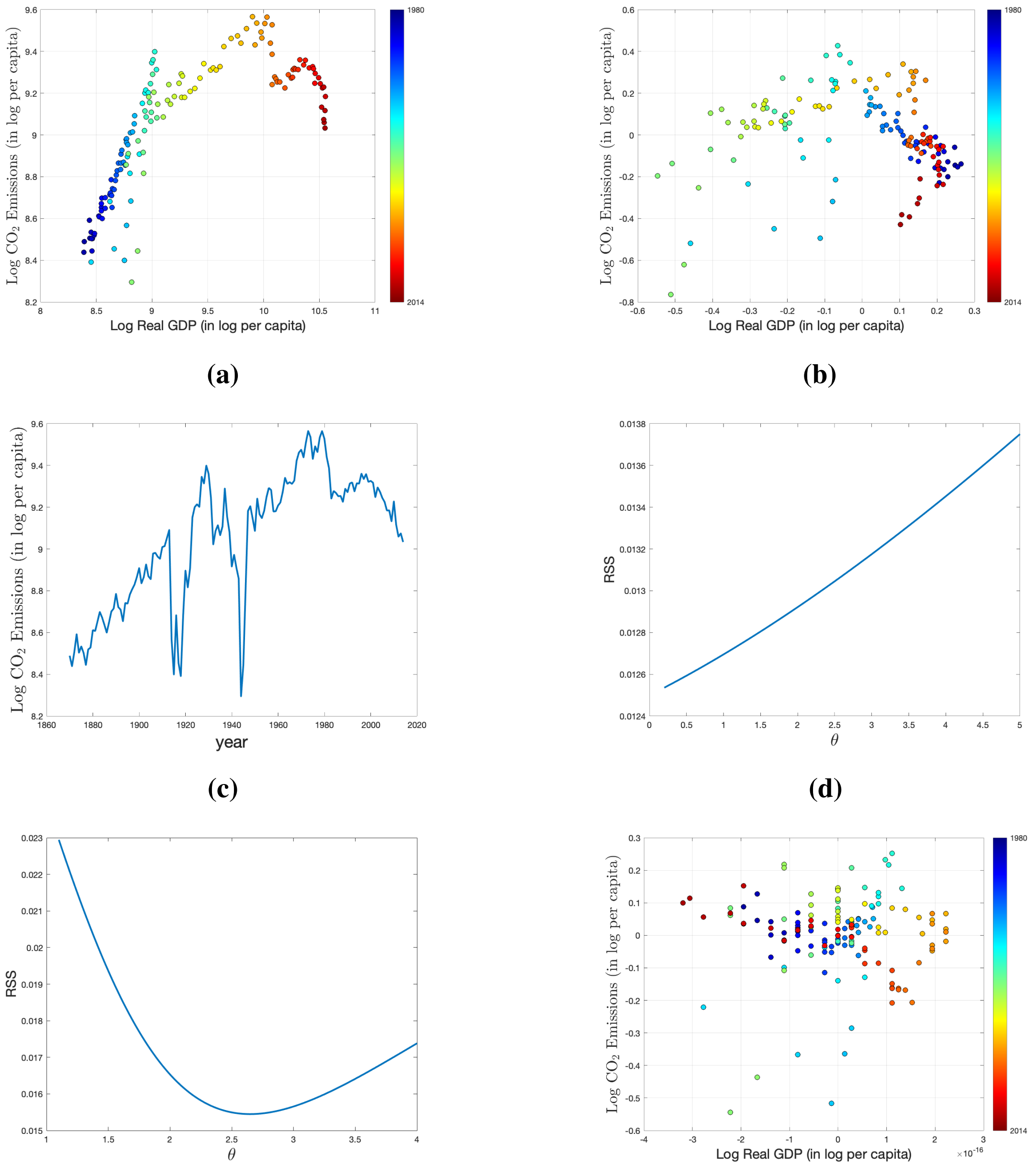


Figure 2: Overview graphs for Belgium over 1870-2014. **(a)** log(GDP) versus log(CO_2) (both per capita). **(b)** As subfigure (a) but using detrended variables. **(c)** The log per capita CO_2 emissions time series for Belgium. **(d)** The residual sum of squares (RSS) for the nonlinear model specification $y_t = \tau_1 + \tau_2 t + \phi_1 x_t + \phi_2 x_t^\theta + u_t$ for various values of θ. **(e)** The RSS as a function of θ for the flexible nonlinear trend specification $y_t = \tau_1 + \tau_2 t + \tau_3 t^\theta + \phi x_t + u_t$. **(f)** The relation between x_t and y_t after partialling out the constant, linear trend, and flexible deterministic trend.

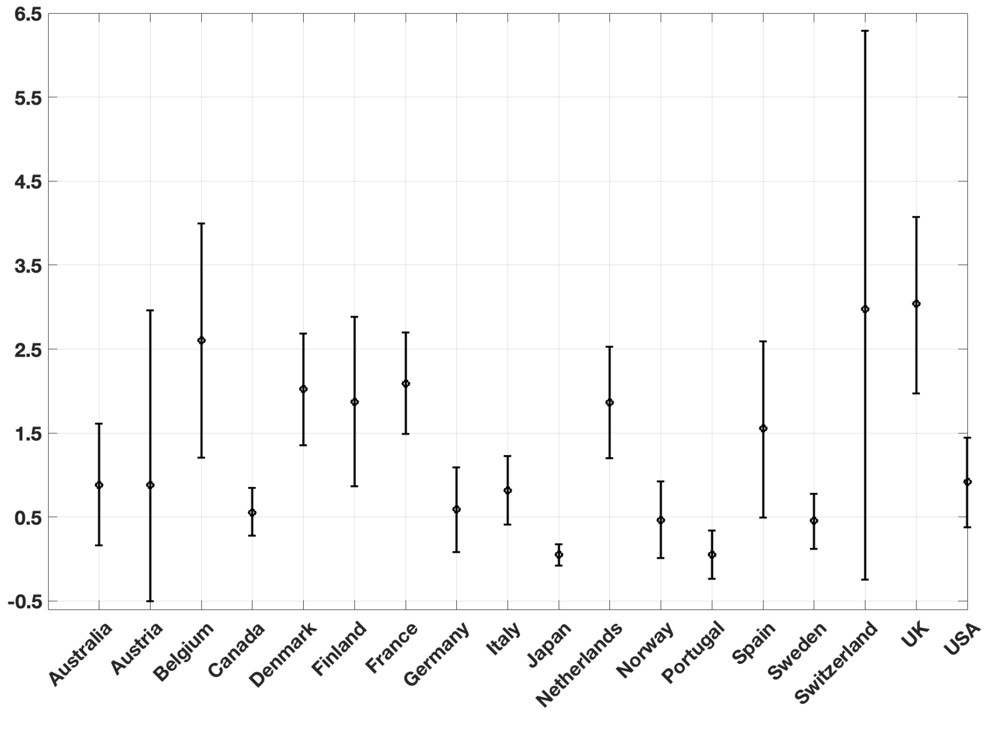


Figure 3: Estimates and 95% confidence intervals for $\widehat{\theta}$ in model specification (M4).

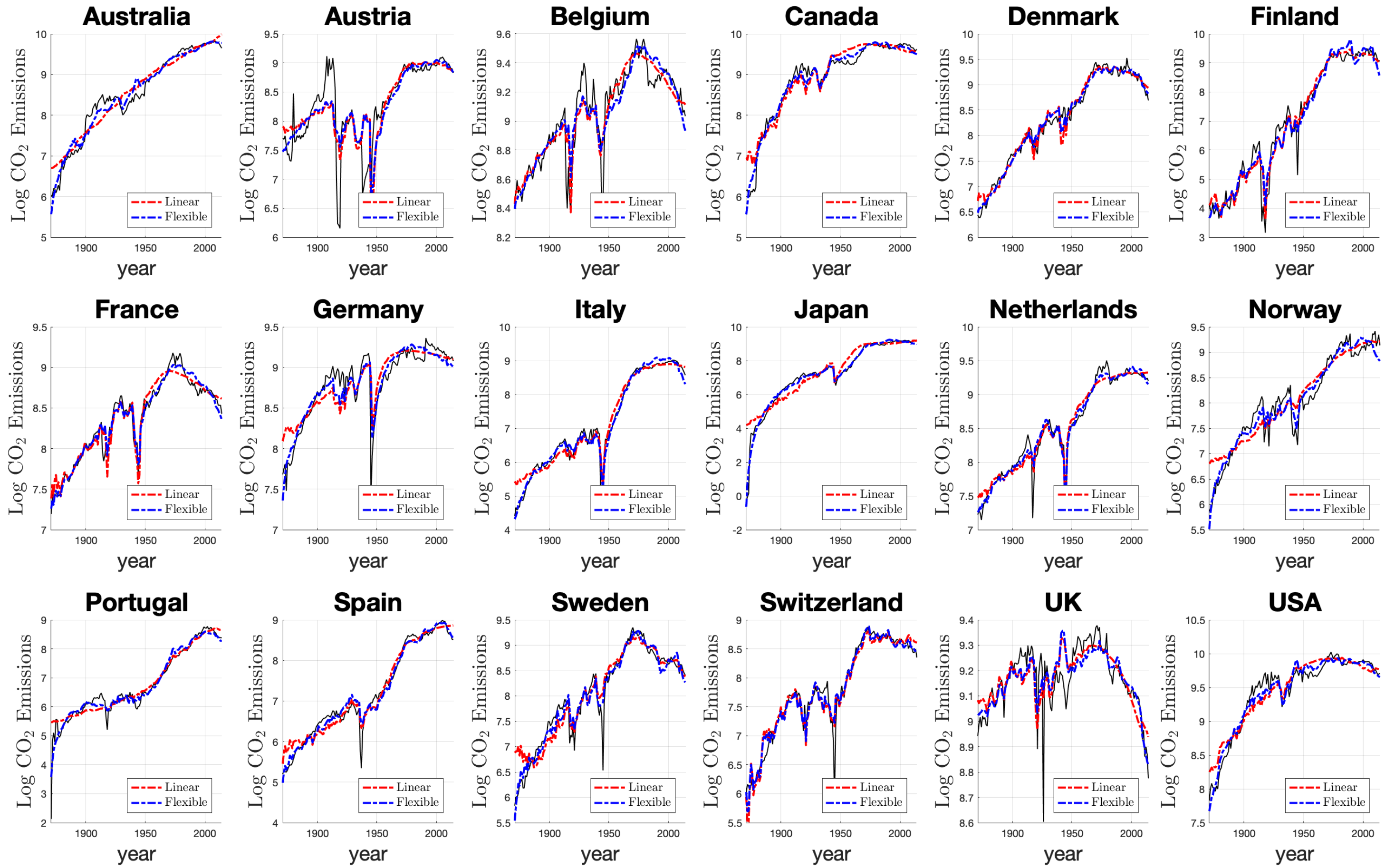


Figure 4: Estimation results for CO_2 emissions: actual values (black), fitted values under the CPR model $y_t = \tau_1 + \tau_2 t + \phi_1 x_t + \phi_2 x_t^2 + u_t$ (red), and fitted values under the GCPR model $y_t = \tau_1 + \tau_2 t + \tau_3 t^\theta + \phi_1 x_t + u_t$ (blue).

Supplemental Appendix to:

Cointegrating Polynomial Regressions with Power Law Trends: A New Angle on the Environmental Kuznets Curve

Yicong Lin and Hanno Reuvers

A Useful Lemmas

In this section, we first show some preliminary results that will be used in the proofs of main theorems (Section B).

Lemma 1

(i) For $a_L > -1$, we have $\sup_{a \geq a_L} \left| \frac{1}{T} \sum_{t=1}^{T} \left(\frac{t}{T}\right)^a \right| \leq C$,

(ii) Under Assumption 2, for any $a > -\frac{1}{2}$, and any $k \geq 0$, $\mathbb{E}\left(\frac{1}{\sqrt{T}} \sum_{t=1}^{T} \left(\frac{t}{T}\right)^a (\ln t)^k u_t\right)^2 \leq C(\ln T)^{2k}$,

(iii) Under Assumption 2, for some a_L and a_U such that $-\frac{1}{2} < a_L < a_U < \infty$, and any $k \geq 0$, $\mathbb{E}\left(\sup_{a \in [a_L, a_U]} \left| \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \left(\frac{t}{T}\right)^a (\ln t)^k u_t \right|\right) \leq C(\ln T)^k$,

(iv) If a_L and a_U satisfy $-1 < a_L < a_U < \infty$, and if $k \in 0, 1, 2, \ldots$, then

$$\sup_{a \in [a_L, a_U]} \left| \frac{1}{T} \sum_{t=1}^{T} \left(\frac{t}{T}\right)^a \left(\ln \frac{t}{T}\right)^k - \int_0^1 r^a (\ln r)^k dr \right| \leq C \frac{(\ln T)^{k+1}}{T^{1+\min(a_L, 0)}}.$$

Proof ***(i)*** This is shown in lemma 4 of Robinson (2012). ***(ii)*** Note that

$$\begin{aligned}
\mathbb{E}\left(\frac{1}{\sqrt{T}} \sum_{t=1}^{T} \left(\frac{t}{T}\right)^a (\ln t)^k u_t\right)^2 &= \frac{1}{T} \sum_{t=1}^{T} \sum_{s=1}^{T} \left(\frac{t}{T}\right)^a \left(\frac{s}{T}\right)^a (\ln t)^k (\ln s)^k \, \mathbb{E}(u_t u_s) \\
&\leq \frac{(\ln T)^{2k}}{T} \sum_{t=1}^{T} \sum_{s=1}^{T} \left(\frac{t}{T}\right)^a \left(\frac{s}{T}\right)^a \left|\mathbb{E}(u_t u_s)\right| \leq 2 \frac{(\ln T)^{2k}}{T} \sum_{t=1}^{T} \sum_{s=0}^{t-1} \left(\frac{t}{T}\right)^a \left(\frac{t-s}{T}\right)^a \left|\gamma_s\right|,
\end{aligned} \tag{A.1}$$

where we define $\gamma_s = \mathbb{E}(u_t u_{t-s})$. For the given index ranges, we also have $|t - s| \leq t$ such that

$$\mathbb{E}\left(\frac{1}{\sqrt{T}} \sum_{t=1}^{T} \left(\frac{t}{T}\right)^a (\ln t)^k u_t\right)^2 \leq 2(\ln T)^{2k} \frac{1}{T} \sum_{t=1}^{T} \left(\frac{t}{T}\right)^{2a} \sum_{s=0}^{\infty} |\gamma_s|. \tag{A.2}$$

The first summation in the RHS of (A.2) is bounded in view of Lemma 1*(i)* and $\sum_{s=0}^{\infty} |\gamma_s| < \infty$ due to Assumption 2(a) (cf. Appendix 3.A. in Hamilton (1994)). ***(iii)*** Using the equality $\frac{t}{T} = \sum_{s=0}^{t-1} \left[\left(\frac{s+1}{T}\right)^a - \left(\frac{s}{T}\right)^a\right]$ and a change in the order of summation, we find

$$\begin{aligned}
\sum_{t=1}^{T} \left(\frac{t}{T}\right)^a (\ln t)^k u_t &= \sum_{t=1}^{T} \sum_{s=0}^{t-1} \left[\left(\frac{s+1}{T}\right)^a - \left(\frac{s}{T}\right)^a\right] (\ln t)^k u_t = \sum_{s=0}^{T-1} \left[\left(\frac{s+1}{T}\right)^a - \left(\frac{s}{T}\right)^a\right] \sum_{t=s+1}^{T} (\ln t)^k u_t \\
&= \left(\frac{1}{T}\right)^a \sum_{t=1}^{T} (\ln t)^k u_t + \sum_{s=1}^{T-1} \left[\left(\frac{s+1}{T}\right)^a - \left(\frac{s}{T}\right)^a\right] \left(\sum_{t=1}^{T} (\ln t)^k u_t - \sum_{t=1}^{s} (\ln t)^k u_t\right) \\
&= \left(\frac{1}{T}\right)^a \sum_{t=1}^{T} (\ln t)^k u_t + \sum_{t=1}^{T} (\ln t)^k u_t - \left(\frac{1}{T}\right)^a \sum_{t=1}^{T} (\ln t)^k u_t - \sum_{s=1}^{T-1} \left[\left(\frac{s+1}{T}\right)^a - \left(\frac{s}{T}\right)^a\right] \sum_{t=1}^{s} (\ln t)^k u_t,
\end{aligned}$$

and hence

$$\mathbb{E}\left(\sup_{a\in[a_L,a_U]}\left|\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left(\frac{t}{T}\right)^a(\ln t)^k u_t\right|\right)\le \mathbb{E}\left|\frac{1}{\sqrt{T}}\sum_{t=1}^{T}(\ln t)^k u_t\right| + \mathbb{E}\left(\sup_{a\in[a_L,a_U]}\left|\frac{1}{\sqrt{T}}\sum_{s=1}^{T-1}\left[\left(\frac{s+1}{T}\right)^a-\left(\frac{s}{T}\right)^a\right]\sum_{t=1}^{s}(\ln t)^k u_t\right|\right). \tag{A.3}$$

For the first term in the RHS of (A.3), we have $\mathbb{E}\left|\frac{1}{\sqrt{T}}\sum_{t=1}^{T}(\ln t)^k u_t\right| \le \left(\mathbb{E}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}(\ln t)^k u_t\right)^2\right)^{1/2} \le C(\ln T)^k$ by Lemma 1(ii) with $a=0$. For the second term, note that

$$\left|\frac{1}{\sqrt{T}}\sum_{s=1}^{T-1}\left[\left(\frac{s+1}{T}\right)^a-\left(\frac{s}{T}\right)^a\right]\sum_{t=1}^{s}(\ln t)^k u_t\right| \le \frac{1}{\sqrt{T}}\sum_{s=1}^{T-1}\left(\frac{s}{T}\right)^a\left|\left(1+\frac{1}{s}\right)^a-1\right|\left|\sum_{t=1}^{s}(\ln t)^k u_t\right|. \tag{A.4}$$

To deal with the supremum of $\left|\left(1+\frac{1}{s}\right)^a-1\right|$ over $[a_L,a_U]$, we define $g_a(x)=(1+x)^a-1$ for $0\le x\le 1$. If $-\frac{1}{2}<a\le 1$, then $|g_a(x)|\le |a|x$ by Bernoulli's inequality. If $a\ge 1$, then convexity of $g_a(x)$ implies

$$g_a(x)\le (1-x)g_a(0)+xg_a(1)\le (2^a-1)\,x.$$

We conclude that $|g_a(x)|\le Cx$ for all $a_L\le a\le a_U$ and $x\in[0,1]$. Combining this result with (A.4), we have

$$\begin{aligned}
&\mathbb{E}\left(\sup_{a\in[a_L,a_U]}\left|\frac{1}{\sqrt{T}}\sum_{s=1}^{T-1}\left[\left(\frac{s+1}{T}\right)^a-\left(\frac{s}{T}\right)^a\right]\sum_{t=1}^{s}(\ln t)^k u_t\right|\right)\\
&\le \mathbb{E}\left(\frac{1}{\sqrt{T}}\sum_{s=1}^{T-1}\left(\frac{s}{T}\right)^{a_L}\sup_{a\in[a_L,a_U]}\left|\left(1+\frac{1}{s}\right)^a-1\right|\left|\sum_{t=1}^{s}(\ln t)^k u_t\right|\right)\le C\frac{1}{\sqrt{T}}\sum_{s=1}^{T-1}\left(\frac{s}{T}\right)^{a_L}\frac{1}{s}\,\mathbb{E}\left|\sum_{t=1}^{s}(\ln t)^k u_t\right|\\
&\le CT^{-(a_L+1/2)}\sum_{s=1}^{T-1}s^{a_L-1/2}(\ln s)^k\le C(\ln T)^k\left[\frac{1}{T}\sum_{s=1}^{T}\left(\frac{s}{T}\right)^{a_L-1/2}\right]\le C(\ln T)^k,
\end{aligned}$$

where we used $\mathbb{E}\left|\sum_{t=1}^{s}(\ln t)^k u_t\right| \le \left(\mathbb{E}\left(\sum_{t=1}^{s}(\ln t)^k u_t\right)^2\right)^{1/2} \le Cs^{1/2}(\ln s)^k$ (the steps in the proof of (ii) require a small modification to establish this) to go to the last line, and (i) to obtain the final inequality. The proof is complete since we have bounded both terms in the RHS of (A.3). ***(iv)*** If we divide the integral into integration intervals of width $\frac{1}{T}$, then we find

$$\begin{aligned}
&\sup_{a\in[a_L,a_U]}\left|\frac{1}{T}\sum_{t=1}^{T}\left(\frac{t}{T}\right)^a\left(\ln\frac{t}{T}\right)^k-\int_0^1 r^a(\ln r)^k dr\right|\\
&=\sup_{a\in[a_L,a_U]}\left|\sum_{t=1}^{T}\int_{(t-1)/T}^{t/T}\left(\frac{t}{T}\right)^a\left(\ln\frac{t}{T}\right)^k dr-\sum_{t=1}^{T}\int_{(t-1)/T}^{t/T}r^a(\ln r)^k dr\right|\\
&=\sup_{a\in[a_L,a_U]}\left|\frac{1}{T}\left(\frac{1}{T}\right)^a\left(\ln\frac{1}{T}\right)^k-\int_0^{1/T}r^a(\ln r)^k dr+\sum_{t=2}^{T}\int_{(t-1)/T}^{t/T}\left[\left(\frac{t}{T}\right)^a\left(\ln\frac{t}{T}\right)^k-r^a(\ln r)^k\right]dr\right| \qquad (A.5)\\
&\le \sup_{a\in[a_L,a_U]}\left|\left(\frac{1}{T}\right)^{a+1}\left(\ln\frac{1}{T}\right)^k\right|+\sup_{a\in[a_L,a_U]}\left|\int_0^{1/T}r^a(\ln r)^k dr\right|\\
&\qquad+\sup_{a\in[a_L,a_U]}\sum_{t=2}^{T}\int_{(t-1)/T}^{t/T}\left|\left(\frac{t}{T}\right)^a\left(\ln\frac{t}{T}\right)^k-r^a(\ln r)^k\right|dr=:Ia+Ib+Ic,
\end{aligned}$$

using the triangle inequality. Clearly, Ia is bounded by $T^{-(a_L+1)}(\ln T)^k$. For Ib we can use the standard integral (cf. Adams and Essex (2016)), namely $\int_0^{1/T} r^a(\ln r)^k dr = \frac{(-1)^k}{a+1}\left(\frac{1}{T}\right)^{a+1}(\ln T)^k - \frac{k}{a+1}\int_0^{1/T} r^a(\ln r)^{k-1}dr$ for $k \neq -1$, to obtain

$$\begin{aligned}\int_0^{1/T} r^a(\ln r)^k dr &= (-1)^k\left(\frac{1}{T}\right)^{a+1}\sum_{j=0}^{k-1}\frac{k!}{(k-j)!}\frac{1}{(a+1)^{1+j}}(\ln T)^{k-j} + (-1)^k\frac{k!}{(a+1)^k}\int_0^{1/T} r^a dr \\ &= (-1)^k\left(\frac{1}{T}\right)^{a+1}\sum_{j=0}^{k}\frac{k!}{(k-j)!}\frac{1}{(a+1)^{1+j}}(\ln T)^{k-j}.\end{aligned}$$

We therefore conclude that

$$\begin{aligned}Ib &\leq \sum_{j=1}^{k}\frac{k!}{(k-j)!}\sup_{a\in[a_L,a_U]}\frac{1}{(a+1)^{1+j}}\left(\frac{1}{T}\right)^{a+1}(\ln T)^{k-j} \leq \sum_{j=1}^{k}\frac{k!}{(k-j)!}\frac{1}{(a_L+1)^{1+j}}\left(\frac{1}{T}\right)^{a_L+1}(\ln T)^{k-j} \\ &\leq CT^{-(a_L+1)}(\ln T)^k.\end{aligned}$$

It remains to bound the term Ic. Changing the integration variable to $r = \frac{t}{T} - s$ yields

$$Ic = \sup_{a\in[a_L,a_U]}\sum_{t=2}^{T}\int_0^{1/T}\left|\left(\frac{t}{T}\right)^a\left(\ln\frac{t}{T}\right)^k - \left(\frac{t}{T}-s\right)^a\left[\ln\left(\frac{t}{T}-s\right)\right]^k\right| ds. \tag{A.6}$$

We subsequently derive an upper bound for the integrand using an approach which mimics the derivations in (D.14) and (D.15) in Robinson (2012). For any $\frac{2}{T} \leq \ell \leq 1$ (such that $0 < s/\ell \leq \frac{1}{2}$), we have

$$\begin{aligned}\left|\ell^a(\ln\ell)^k - (\ell-s)^a(\ln(\ell-s))^k\right| &= \left|[\ell^a - (\ell-s)^a](\ln\ell)^k + (\ell-s)^a[(\ln\ell)^k - (\ln(\ell-s))^k]\right| \\ &\leq \left|[\ell^a - (\ell-s)^a](\ln\ell)^k\right| + \left|(\ell-s)^a\left[(\ln\ell)^k - (\ln(\ell-s))^k\right]\right| \\ &= \ell^a\left|\ln\ell\right|^k\left|1-(1-s/\ell)^a\right| + \ell^a(1-s/\ell)^a\left|(\ln\ell)^k - (\ln(\ell-s))^k\right| =: IIa + IIb,\end{aligned} \tag{A.7}$$

by the triangle inequality and the fact that $\left|(\ell-s)^a\right| = (\ell-s)^a$. For IIa similar arguments as those found below (A.4) give $\left|1-(1-x)^a\right| \leq Cx$, and hence

$$IIa \leq C\ell^{a_L}\left|\ln\ell\right|^k\frac{s}{\ell} \leq C\ell^{a_L-1}\left|\ln\ell\right|^k s \leq C\ell^{a_L-1}\left|\ln\ell\right|^k\frac{1}{T} \leq C\ell^{a_L-1}\left|\ln\ell\right|^k\frac{1}{T} \leq C\ell^{a_L-1}(\ln T)^k\frac{1}{T}, \tag{A.8}$$

since $|\ln\ell| \leq |\ln T|$ for all $\frac{2}{T} \leq \ell \leq 1$. For IIb we first note that $\frac{1}{2} \leq 1 - s/\ell < 1$ and therefore $(1-s/\ell)^a < (1-s/\ell)^{-1} \leq 2$. Moreover, we use the factorization $p^n - q^n = (p-q)\sum_{j=0}^{n-1}p^{n-1-j}q^j$ to obtain[16]

$$\begin{aligned}\left|(\ln\ell)^k - (\ln(\ell-s))^k\right| &= \left|\ln\ell - \ln(\ell-s)\right|\left|\sum_{j=0}^{k-1}(\ln\ell)^{k-1-j}(\ln(\ell-s))^j\right| \\ &= \left|\ln(1-s/\ell)\right|\left|\sum_{j=0}^{k-1}(\ln\ell)^{k-1-j}(\ln(\ell-s))^j\right| \leq \left|\ln(1-s/\ell)\right|\sum_{j=0}^{k-1}|\ln\ell|^{k-1-j}|\ln(\ell-s)|^j \\ &\leq k\,|\ln(1-s/\ell)|\,(\ln T)^{k-1} \leq 2k\frac{s}{\ell}(\ln T)^{k-1},\end{aligned} \tag{A.9}$$

[16]For any $x > -1$, we have the inequality $\frac{x}{1+x} \leq \ln(1+x) \leq x$. This implies that $|\ln(1-s/\ell)| = -\ln(1-s/\ell) \leq \frac{s/\ell}{1-s/\ell} \leq 2\frac{s}{\ell}$.

because $1/T \le \ell - s < 1$ and thus $|\ln(\ell - s)| \le \ln T$. Combining all previous results for IIb gives

$$IIb \le C\ell^{a}\frac{s}{\ell}(\ln T)^{k-1} \le C\ell^{a_L-1}(\ln T)^{k-1}\frac{1}{T}.$$

Since $\frac{2}{T} \le \ell \le 1$, we use the bounds on IIa and IIb to bound the integrand of (A.6) as follows:

$$Ic \le C \sup_{a\in[a_L,a_U]} \sum_{t=2}^{T} \int_0^{1/T} \left(\frac{t}{T}\right)^{a_L-1} \frac{1}{T}(\ln T)^k ds \le C\frac{(\ln T)^k}{T^2} \sum_{t=1}^{T} \left(\frac{t}{T}\right)^{a_L-1}.$$

The asymptotic order of $\sum_{t=1}^{T}(\frac{t}{T})^{a_L-1}$ relies on the values of a_L. We distinguish three cases: (1) if $a_L < 0$, then $\sum_{t=1}^{T}(\frac{t}{T})^{a_L-1} = T^{1-a_L}\sum_{t=1}^{T}\frac{1}{t^{1-a_L}} = T^{1-a_L}O(1)$, (2) if $a_L = 0$, then $\sum_{t=1}^{T}(\frac{t}{T})^{a_L-1} = T\sum_{t=1}^{T} t^{-1} = TO(\ln T)$, and (3) if $a_L > 0$, $\sum_{t=1}^{T}(\frac{t}{T})^{a_L-1} = O(T)$ by Lemma 1(i). Overall, we have

$$Ic \le C\frac{(\ln T)^k}{T^2}\sum_{t=1}^{T}\left(\frac{t}{T}\right)^{a_l-1} = O\left(\frac{(\ln T)^k}{T^{a_l+1}}\mathbb{1}_{\{a_l<0\}} + \frac{(\ln T)^{k+1}}{T}\mathbb{1}_{\{a_l=0\}} + \frac{(\ln T)^k}{T}\mathbb{1}_{\{a_l>0\}}\right). \tag{A.10}$$

It is seen that Ia, Ib, and Ic converge to zero as $T \to \infty$. The proof follows from (A.5). ■

Lemma 2

Let Assumption 2 hold. For any a such that $-\frac{1}{2} < a_L \le a \le a_U < \infty$, any $i \in \{1,2,\dots,m\}$, any $j \in \{1,2,\dots,p_i\}$, and $k \in \{0,1,2,\dots\}$, we have:

(i) $\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left(\frac{x_{it}}{\sqrt{T}}\right)^j u_t \Rightarrow \int_0^1 \boldsymbol{B}_{v_i}^j(r)dB_u(r) + j\boldsymbol{\Delta}_{v_iu}\int_0^1 \boldsymbol{B}_{v_i}^{j-1}(r)dr,$

(ii) $\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left(\frac{t}{T}\right)^a\left(\ln\frac{t}{T}\right)^k u_t \Rightarrow \int_0^1 r^a(\ln r)^k dB_u(r),$

(iii) $\frac{1}{T}\sum_{t=1}^{T}\left(\frac{t}{T}\right)^a\left(\ln\frac{t}{T}\right)^k\left(\frac{x_{it}}{\sqrt{T}}\right)^j \Rightarrow \int_0^1 r^a(\ln r)^k\boldsymbol{B}_{v_i}^j(r)dr.$

Moreover, the weak convergence in (i)-(iii) holds jointly.

Proof For $r \in (0,1]$, we define $f(r) = r^a(\ln r)^k$. Two partial sum processes are defined as $S_T(r) = \frac{1}{\sqrt{T}}\sum_{s=1}^{[rT]} u_s$, and $X_{i,T}(r) = \frac{1}{\sqrt{T}}x_{i,[rT]} = \frac{1}{\sqrt{T}}\sum_{i=1}^{[rT]} v_{it}$. Finally, set $f_T(r) = \left(\frac{[rT]}{T}\right)^a\left(\ln\frac{[rT]}{T}\right)^k$ for $r \in \left[\frac{1}{T},1\right]$.
(i) This result follows from lemma 1 of Hong and Phillips (2010). ***(ii)*** We have

$$\begin{aligned}\frac{1}{\sqrt{T}}\left(\frac{t}{T}\right)^a\left(\ln\frac{t}{T}\right)^k u_t &= f_T\left(\frac{t}{T}\right)\frac{u_t}{\sqrt{T}} = f_T\left(\frac{t}{T}\right)\left[S_T\left(\frac{t}{T}\right) - S_T\left(\frac{t-1}{T}\right)\right]\\ &= \left[f_T\left(\frac{t}{T}\right)S_T\left(\frac{t}{T}\right) - f_T\left(\frac{t-1}{T}\right)S_T\left(\frac{t-1}{T}\right)\right] - \left[f_T\left(\frac{t}{T}\right) - f_T\left(\frac{t-1}{T}\right)\right]S_T\left(\frac{t-1}{T}\right)\end{aligned} \tag{A.11}$$

and hence

$$\begin{aligned}\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left(\frac{t}{T}\right)^a\left(\ln\frac{t}{T}\right)^k u_t &= \left(\frac{1}{T}\right)^a\left(\ln\frac{1}{T}\right)^k\frac{u_1}{\sqrt{T}} + \frac{1}{\sqrt{T}}\sum_{t=2}^{T}\left(\frac{t}{T}\right)^a\left(\ln\frac{t}{T}\right)^k u_t\\ &\overset{(A.11)}{=} f_T\left(\frac{1}{T}\right)S_T\left(\frac{1}{T}\right) + \left[f_T(1)S_T(1) - f_T\left(\frac{1}{T}\right)S_T\left(\frac{1}{T}\right)\right] - \sum_{t=2}^{T}\left[f_T\left(\frac{t}{T}\right) - f_T\left(\frac{t-1}{T}\right)\right]S_T\left(\frac{t-1}{T}\right)\\ &\overset{f_T(1)=0}{=} -\sum_{t=2}^{T}\int_{(t-1)/T}^{t/T} S_T(r)df_T(r)\end{aligned} \tag{A.12}$$

where we used the fact that $S_T(\cdot)$ is piecewise constant. In view of Assumption 2, we can extend suitably extend the probability space and have the following uniformly strong approximation of the partial sum process S_T (see for example page 562 of Phillips (2007)):

$$\sup_{1\le t\le T}\left|S_T\left(\frac{t-1}{T}\right)-B_u\left(\frac{t-1}{T}\right)\right| = o_{a.s.}\left(\frac{1}{T^{(1/2)-(1/q)}}\right), \tag{A.13}$$

for $q>2$. Continuing from (A.12), this uniformly strong approximation gives

$$\begin{aligned}
\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left(\frac{t}{T}\right)^a\left(\ln\frac{t}{T}\right)^k u_t &= -\sum_{t=2}^{T}\int_{(t-1)/T}^{t/T} B_u(r)df_T(r) + o_{a.s.}\left(\frac{1}{T^{(1/2)-(1/q)}}\right)\\
&= -\int_{1/T}^{1} B_u(r)df_T(r) + o_{a.s.}\left(\frac{1}{T^{(1/2)-(1/q)}}\right)\\
&= B_u\left(\frac{1}{T}\right)f_T\left(\frac{1}{T}\right) + \int_{1/T}^{1} f_T(r)dB_u(r) + o_{a.s.}\left(\frac{1}{T^{(1/2)-(1/p)}}\right)\\
&= \int_0^1 f(r)dB_u(r) - \int_0^{1/T} f(r)dB_u(r) + B_u\left(\frac{1}{T}\right)f_T\left(\frac{1}{T}\right)\\
&\quad + \int_{1/T}^{1}\Big[f_T(r)-f(r)\Big]dB_u(r) + o_{a.s.}\left(\frac{1}{T^{(1/2)-(1/p)}}\right),
\end{aligned} \tag{A.14}$$

where the third line is obtained using integration by parts of the mean square Riemann-Stieltjes integral, c.f. theorem 2.7 in Tanaka (2017). It remains to show that $\int_0^{1/T} f(r)dB_u(r)$, $B_u\left(\frac{1}{T}\right)f_T\left(\frac{1}{T}\right)$, and $\int_{1/T}^1\Big[f_T(r)-f(r)\Big]dB_u(r)$ are asymptotically negligible. These quantities are zero mean so it suffices to show that their variances vanish as $T\to\infty$. By the isometry property and steps similar to those above (A.6), we have

$$\mathbb{V}\mathrm{ar}\left(\int_0^{1/T} f(r)dB_u(r)\right) = \Omega_{uu}\int_0^{1/T}[f(r)]^2dr \le CT^{-(2a_L+1)}(\ln T)^{2k}\to 0, \tag{A.15}$$

as $T\to\infty$. Also, $\mathbb{V}\mathrm{ar}\left(B_u\left(\frac{1}{T}\right)f_T\left(\frac{1}{T}\right)\right) = \frac{1}{T}\Omega_{uu}\left[f_T\left(\frac{1}{T}\right)\right]^2 = \Omega_{uu}\left(\frac{1}{T}\right)^{2a_L+1}\left(\ln\frac{1}{T}\right)^{2k}\to 0$. To control the variance of $\int_{1/T}^1\Big[f_T(r)-f(r)\Big]dB_u(r)$, we look at

$$\begin{aligned}
\int_{1/T}^{1}|f(r)-f_T(r)|^2\,dr &= \sum_{t=2}^{T}\int_{(t-1)/T}^{t/T}\left|f(r)-\left(\frac{t-1}{T}\right)^a\left(\ln\frac{t-1}{T}\right)^k\right|^2 dr\\
&= \sum_{t=1}^{T-1}\int_{t/T}^{(t+1)/T}\left|r^a(\ln r)^k-\left(\frac{t}{T}\right)^a\left(\ln\frac{t}{T}\right)^k\right|^2 dr\\
&= \sum_{t=1}^{T-1}\int_0^{1/T}\left|\left(\frac{t}{T}+s\right)^a\left[\ln\left(\frac{t}{T}+s\right)\right]^k-\left(\frac{t}{T}\right)^a\left(\ln\frac{t}{T}\right)^k\right|^2 ds.
\end{aligned} \tag{A.16}$$

Now let $\ell\in\left\{\frac{1}{T},\frac{2}{T},\dots,1\right\}$ and recall that $0\le s\le\frac{1}{T}$ (hence also $0\le\frac{s}{\ell}\le 1$). Using the triangle inequality, the expression in absolute values can be bounded as

$$\begin{aligned}
\left|(\ell+s)^a(\ln(\ell+s))^k-\ell^a(\ln\ell)^k\right| &= \left|[(\ell+s)^a-\ell^a](\ln(\ell+s))^k+\ell^a\left[(\ln(\ell+s))^k-(\ln\ell)^k\right]\right|\\
&\le \left|[(\ell+s)^a-\ell^a](\ln(\ell+s))^k\right| + \left|\ell^a\left[(\ln(\ell+s))^k-(\ln\ell)^k\right]\right|\\
&= \ell^a\left|(1+s/\ell)^a-1\right||\ln(\ell+s)|^k + \ell^a\left|(\ln(\ell+s))^k-(\ln\ell)^k\right| = IIc+IId.
\end{aligned} \tag{A.17}$$

By the inequality $|g_a(x)| \leq Cx$ below (A.4) and the fact that $|\ln(\ell + s)| \leq |\ln \ell| + |\ln(1 + s/\ell)| \leq \ln T + s/\ell$, we obtain $IIc \leq C\ell^{a_L} \frac{s}{\ell} \left|\ln T + \frac{s}{\ell}\right|^k \leq C\ell^{a_L - 1}(\ln T)^k \frac{1}{T}$. Moreover, the factorisation $p^n - q^n = (p - q) \sum_{j=0}^{n-1} p^{n-1-j} q^j$ yields

$$IId = \ell^a |\ln(1 + s/\ell)| \left| \sum_{j=0}^{k-1} (\ln(\ell + s))^{k-1-j} (\ln \ell)^j \right| \leq k\ell^{a_L} \frac{s}{\ell} |(\ln T) + 1|^{k-1} \leq C\ell^{a_L - 1}(\ln T)^{k-1} \frac{1}{T}. \tag{A.18}$$

By combination of the bounds on IIc and IId, we conclude that $\left|(\ell + s)^a(\ln(\ell + s))^k - \ell^a(\ln \ell)^k\right| \leq C\ell^{a_L - 1}(\ln T)^k \frac{1}{T}$ and arrive at the following upper bound on the RHS of (A.16):

$$\begin{aligned} \int_{1/T}^1 |f(r) - f_T(r)|^2 \, dr &\leq C(\ln T)^{2k} \frac{1}{T^3} \sum_{t=1}^T \left(\frac{t}{T}\right)^{2(a_L - 1)} \\ &= O\left(\frac{(\ln T)^{2k}}{T^{2(a_L + \frac{1}{2})}} \mathbb{1}_{\{a_L < \frac{1}{2}\}} + \frac{(\ln T)^{2k+1}}{T^2} \mathbb{1}_{\{a_L = \frac{1}{2}\}} + \frac{(\ln T)^{2k}}{T^2} \mathbb{1}_{\{a_L > \frac{1}{2}\}} \right). \end{aligned} \tag{A.19}$$

The RHS of (A.19) will go to zero as $T \to \infty$, thereby establishing that $\int_{1/T}^1 \left[f_T(r) - f(r)\right] dB_u(r)$ is also asymptotically negligible. The proof of part (ii) is now complete. ***(iii)*** We have

$$\begin{aligned} \frac{1}{T} \sum_{t=1}^T \left(\frac{t}{T}\right)^a \left(\ln \frac{t}{T}\right)^k \left(\frac{x_{it}}{\sqrt{T}}\right)^j &= \sum_{t=2}^T \int_{(t-1)/T}^{t/T} f_T(r) X_{i,T}^j(r) dt \\ &= \int_0^1 f(r) X_{i,T}^j(r) dr + \int_{1/T}^1 \left[f_T(r) - f(r)\right] X_{i,T}^j(r) dr =: IIIa + IIIb. \end{aligned} \tag{A.20}$$

Given the CMT and $X_{i,T} \Rightarrow B_{v_i}$, term $IIIa$ will converge weakly to $\int_0^1 f(r) \boldsymbol{B}_{v_i}^j(r) dr$ if we can show that $x \mapsto \int_0^1 f(r) x^j(r) dr$ is a continuous functional. Let $x, y \in D[0, 1]$. Hölder's inequality implies

$$\begin{aligned} \left| \int_0^1 f(r) x^j(r) dr - \int_0^1 f(r) y^j(r) dr \right| &= \left| \int_0^1 f(r) \left(x^j(r) - y^j(r)\right) dr \right| \\ &\leq \int_0^1 |f(r)| dr \sup_{r \in [0,1]} |x^j(r) - y^j(r)| \leq C \sup_{r \in [0,1]} |x(r) - y(r)| \to 0, \end{aligned} \tag{A.21}$$

because $\int_0^1 |f(r)| dr = \frac{k!}{(1+a)^{k+1}}$ is bounded. Continuity of the functional now follows from (A.21). If we apply the Cauchy-Schwartz inequality to $IIIb$, then we find

$$IIIb \leq \left[\int_{1/T}^1 |f(r) - f_T(r)|^2 \, dr \right]^{1/2} \left[\int_{1/T}^1 X_{i,T}^{2j}(r) dr \right]^{1/2}.$$

Since $\int_{1/T}^1 |f(r) - f_T(r)|^2 = o(1)$ by (A.19) and $\int_{1/T}^1 X_{i,T}^{2j}(r) dr = \int X_{i,T}^{2j}(r) dr \Rightarrow \int B_{v_i}^{2j}(r) dr$. We conclude that $IIIb = o_p(1)$. Now combine the limiting results for $IIIa$ and $IIIb$ to complete the argument. ■

Lemma 3

Let $f(\boldsymbol{w}_t, \boldsymbol{\gamma}) = \boldsymbol{d}_t(\boldsymbol{\theta})' \boldsymbol{\tau} + \boldsymbol{s}_t' \boldsymbol{\phi}$, where $\boldsymbol{w}_t = [t, \boldsymbol{x}_t]'$ and $\boldsymbol{x}_t = [x_{1t}, x_{2t}, \ldots, x_{mt}]'$ stacks all the stochastic trends. If $\dot{f}(\boldsymbol{w}_t, \boldsymbol{\gamma})$ and $\ddot{f}(\boldsymbol{w}_t, \boldsymbol{\gamma})$ denote the first and second derivatives of $f(\boldsymbol{w}_t, \boldsymbol{\gamma})$ with respect to $\boldsymbol{\gamma}$, then *(i)* $\dot{f}(\boldsymbol{w}_t, \gamma) = [(\boldsymbol{\tau} \odot \boldsymbol{d}_t(\boldsymbol{\theta}))' \ln t, \boldsymbol{d}_t(\boldsymbol{\theta})', \boldsymbol{s}_t']'$,

(ii) $\boldsymbol{L}_{\boldsymbol{\tau}_0,T}\dot{f}(\boldsymbol{w}_t, \gamma) = \begin{bmatrix} [(\boldsymbol{\tau}-\boldsymbol{\tau}_0)\odot\boldsymbol{d}_t(\boldsymbol{\theta})]\ln t + [\boldsymbol{\tau}_0\odot\boldsymbol{d}_t(\boldsymbol{\theta})]\ln\frac{t}{T} \\ \boldsymbol{d}_t(\boldsymbol{\theta}) \\ \boldsymbol{s}_t \end{bmatrix}$, *and hence* $\boldsymbol{L}_{\boldsymbol{\tau}_0,T}\dot{f}(\boldsymbol{w}_t, \gamma_0) = \begin{bmatrix} [\boldsymbol{\tau}_0\odot\boldsymbol{d}_t(\boldsymbol{\theta}_0)]\ln\frac{t}{T} \\ \boldsymbol{d}_t(\boldsymbol{\theta}_0) \\ \boldsymbol{s}_t \end{bmatrix}$,

(iii) *We have*

$$\dot{f}(\boldsymbol{w}_t, \gamma) - \dot{f}(\boldsymbol{w}_t, \gamma_0) = \begin{bmatrix} \Big((\boldsymbol{\tau}-\boldsymbol{\tau}_0)\odot\big(\boldsymbol{d}_t(\boldsymbol{\theta})-\boldsymbol{d}_t(\boldsymbol{\theta}_0)\big)+\boldsymbol{\tau}_0\odot\big(\boldsymbol{d}_t(\boldsymbol{\theta})-\boldsymbol{d}_t(\boldsymbol{\theta}_0)\big)+(\boldsymbol{\tau}-\boldsymbol{\tau}_0)\odot\boldsymbol{d}_t(\boldsymbol{\theta}_0)\Big)\ln t \\ \boldsymbol{d}_t(\boldsymbol{\theta})-\boldsymbol{d}_t(\boldsymbol{\theta}_0) \\ \mathbf{0}_{p\times 1} \end{bmatrix},$$

which implies

$$\boldsymbol{L}_{\boldsymbol{\tau}_0,T}\left(\dot{f}(\boldsymbol{w}_t, \gamma) - \dot{f}(\boldsymbol{w}_t, \gamma_0)\right) = \begin{bmatrix} \Big((\boldsymbol{\tau}-\boldsymbol{\tau}_0)\odot\big(\boldsymbol{d}_t(\boldsymbol{\theta})-\boldsymbol{d}_t(\boldsymbol{\theta}_0)\big)+(\boldsymbol{\tau}-\boldsymbol{\tau}_0)\odot\boldsymbol{d}_t(\boldsymbol{\theta}_0)\Big)\ln t+\boldsymbol{\tau}_0\odot\big(\boldsymbol{d}_t(\boldsymbol{\theta})-\boldsymbol{d}_t(\boldsymbol{\theta}_0)\big)\ln\frac{t}{T} \\ \boldsymbol{d}_t(\boldsymbol{\theta})-\boldsymbol{d}_t(\boldsymbol{\theta}_0) \\ \mathbf{0}_{p\times 1} \end{bmatrix}. \tag{A.22}$$

(iv) $\ddot{f}(\boldsymbol{w}_t, \gamma) = \begin{bmatrix} \operatorname{diag}[\boldsymbol{\tau}]\operatorname{diag}[\boldsymbol{d}_t(\boldsymbol{\theta})](\ln t)^2 & \operatorname{diag}[\boldsymbol{d}_t(\boldsymbol{\theta})]\ln t & \mathbf{O}_{d\times p} \\ \operatorname{diag}[\boldsymbol{d}_t(\boldsymbol{\theta})]\ln t & \mathbf{O}_{d\times d} & \mathbf{O}_{d\times p} \\ \mathbf{O}_{p\times d} & \mathbf{O}_{p\times d} & \mathbf{O}_{p\times p} \end{bmatrix}$,

(v) $\boldsymbol{G}'^{-1}_{\gamma_0,T}\ddot{f}(\boldsymbol{w}_t, \gamma)\boldsymbol{G}^{-1}_{\gamma_0,T} = \begin{bmatrix} \ddot{\boldsymbol{F}}_{11,t} & \ddot{\boldsymbol{F}}_{12,t} & \mathbf{O}_{d\times p} \\ \ddot{\boldsymbol{F}}_{21,t} & \mathbf{O}_{d\times d} & \mathbf{O}_{d\times p} \\ \mathbf{O}_{p\times d} & \mathbf{O}_{p\times d} & \mathbf{O}_{p\times p} \end{bmatrix}$, *where the blocks in this symmetric matrix are given by:*

$$\ddot{\boldsymbol{F}}_{11,t} = \frac{1}{T}\boldsymbol{D}_{d,T}(\boldsymbol{\theta}_0)^{-1}\left(\operatorname{diag}[\boldsymbol{\tau}]\operatorname{diag}[\boldsymbol{d}_t(\boldsymbol{\theta})](\ln t)^2 - 2\operatorname{diag}[\boldsymbol{\tau}_0]\operatorname{diag}[\boldsymbol{d}_t(\boldsymbol{\theta})]\ln t\ln T\right)\boldsymbol{D}_{d,T}(\boldsymbol{\theta}_0)^{-1}$$
$$\ddot{\boldsymbol{F}}_{12,t} = \ddot{\boldsymbol{F}}_{21,t} = \frac{1}{T}\boldsymbol{D}_{d,T}(\boldsymbol{\theta}_0)^{-1}\operatorname{diag}[\boldsymbol{d}_t(\boldsymbol{\theta})]\ln t\boldsymbol{D}_{d,T}(\boldsymbol{\theta}_0)^{-1}.$$

(vi) *Define a symmetric block matrix* $\boldsymbol{M}_T = [\boldsymbol{M}_{T,ij}]_{1\le i,j\le 3} = \boldsymbol{G}'^{-1}_{\gamma_0,T}\left[\sum_{t=1}^T \dot{f}(\boldsymbol{w}_t, \gamma_0)\dot{f}(\boldsymbol{w}_t, \gamma_0)'\right]\boldsymbol{G}^{-1}_{\gamma_0,T}$ *and a stacked vector* $\boldsymbol{z}_T = [\boldsymbol{z}_{T,i}]_{1\le i\le 3} = \boldsymbol{G}'^{-1}_{\gamma_0,T}\left[\sum_{t=1}^T \dot{f}(\boldsymbol{w}_t, \gamma_0)u_t\right]$, *where* $\boldsymbol{M}_{T,ij} = \boldsymbol{M}'_{T,ji}$, *and*

$$\begin{aligned}
\boldsymbol{M}_{T,11} &= \frac{1}{T}\sum_{t=1}^T\Big(\boldsymbol{\tau}_0\odot\boldsymbol{D}_{d,T}(\boldsymbol{\theta}_0)^{-1}\boldsymbol{d}_t(\boldsymbol{\theta}_0)\Big)\Big(\boldsymbol{\tau}_0\odot\boldsymbol{D}_{d,T}(\boldsymbol{\theta}_0)^{-1}\boldsymbol{d}_t(\boldsymbol{\theta}_0)\Big)'\left(\ln\frac{t}{T}\right)^2 \\
\boldsymbol{M}_{T,12} &= \frac{1}{T}\sum_{t=1}^T\Big(\boldsymbol{\tau}_0\odot\boldsymbol{D}_{d,T}(\boldsymbol{\theta}_0)^{-1}\boldsymbol{d}_t(\boldsymbol{\theta}_0)\Big)[\boldsymbol{D}_{d,T}(\boldsymbol{\theta}_0)^{-1}\boldsymbol{d}_t(\boldsymbol{\theta}_0)]'\ln\frac{t}{T} \\
\boldsymbol{M}_{T,13} &= \frac{1}{T}\sum_{t=1}^T\Big(\boldsymbol{\tau}_0\odot\boldsymbol{D}_{d,T}(\boldsymbol{\theta}_0)^{-1}\boldsymbol{d}_t(\boldsymbol{\theta}_0)\Big)(\boldsymbol{D}_{s,T}^{-1}\boldsymbol{s}_t)'\ln\frac{t}{T} \\
\boldsymbol{M}_{T,22} &= \frac{1}{T}\sum_{t=1}^T[\boldsymbol{D}_{d,T}(\boldsymbol{\theta}_0)^{-1}\boldsymbol{d}_t(\boldsymbol{\theta}_0)][\boldsymbol{D}_{d,T}(\boldsymbol{\theta}_0)^{-1}\boldsymbol{d}_t(\boldsymbol{\theta}_0)]' \\
\boldsymbol{M}_{T,23} &= \frac{1}{T}\sum_{t=1}^T[\boldsymbol{D}_{d,T}(\boldsymbol{\theta}_0)^{-1}\boldsymbol{d}_t(\boldsymbol{\theta}_0)](\boldsymbol{D}_{s,T}^{-1}\boldsymbol{s}_t)' \\
\boldsymbol{M}_{T,33} &= \frac{1}{T}\sum_{t=1}^T(\boldsymbol{D}_{s,T}^{-1}\boldsymbol{s}_t)(\boldsymbol{D}_{s,T}^{-1}\boldsymbol{s}_t)',
\end{aligned} \tag{A.23}$$

moreover,

$$
\begin{aligned}
\boldsymbol{z}_{T,1} &= \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \Big(\boldsymbol{\tau}_0 \odot \boldsymbol{D}_{d,T}(\boldsymbol{\theta}_0)^{-1} \boldsymbol{d}_t(\boldsymbol{\theta}_0)\Big)\Big(\ln \frac{t}{T}\Big) u_t \\
\boldsymbol{z}_{T,2} &= \frac{1}{\sqrt{T}} \sum_{t=1}^{T} [\boldsymbol{D}_{d,T}(\boldsymbol{\theta}_0)^{-1} \boldsymbol{d}_t(\boldsymbol{\theta}_0)] u_t \qquad \text{(A.24)} \\
\boldsymbol{z}_{T,3} &= \frac{1}{\sqrt{T}} \sum_{t=1}^{T} (\boldsymbol{D}_{s,T}^{-1} \boldsymbol{s}_t) u_t.
\end{aligned}
$$

Proof All results follow from linearity of the Hadamard product. ■

Lemma 4
For a constant $\delta > 0$, we define

$$
\mathcal{N}_{\delta,T}(\gamma_0) = \Big\{\gamma \in \Gamma : \ \big\|\boldsymbol{D}_{d,T}(\boldsymbol{\theta}_0)\,(\boldsymbol{\theta} - \boldsymbol{\theta}_0)\big\| < \delta T^{-1/2} \ln T, \big\|\boldsymbol{D}_{s,T}\,(\boldsymbol{\phi} - \boldsymbol{\phi}_0)\big\| < \delta T^{-1/2} \ln T, \\
\Big\|\boldsymbol{D}_{d,T}(\boldsymbol{\theta}_0)\Big((\boldsymbol{\tau} - \boldsymbol{\tau}_0) + [\boldsymbol{\tau}_0 \odot (\boldsymbol{\theta} - \boldsymbol{\theta}_0)] \ln T\Big)\Big\| < \delta T^{-1/2} \ln T\Big\}, \quad \text{(A.25)}
$$

where $\boldsymbol{\gamma}_0 \in \Gamma$ is fixed. Under Assumption 2, for all $k \geq 0$,

(i) $\sup_{\gamma \in \mathcal{N}_{\delta,T}(\gamma_0)} \frac{1}{T} (\ln T)^k \sum_{t=1}^{T} \Big\|\boldsymbol{D}_{d,T}(\boldsymbol{\theta}_0)^{-1}\Big(\boldsymbol{d}_t(\boldsymbol{\theta}) - \boldsymbol{d}_t(\boldsymbol{\theta}_0)\Big)\Big\|^2 = o(1),$

(ii) $\sup_{\gamma \in \mathcal{N}_{\delta,T}(\gamma_0)} \frac{1}{T} (\ln T)^k \sum_{t=1}^{T} \Big\|\boldsymbol{D}_{d,T}(\boldsymbol{\theta}_0)^{-1}\Big[\boldsymbol{\tau}_0 \odot \Big(\boldsymbol{d}_t(\boldsymbol{\theta}) - \boldsymbol{d}_t(\boldsymbol{\theta}_0)\Big)\Big] \ln \frac{t}{T}\Big\|^2 = o(1),$

(iii) $\sup_{\gamma \in \mathcal{N}_{\delta,T}(\gamma_0)} \frac{1}{T} (\ln T)^k \sum_{t=1}^{T} \Big\|\boldsymbol{D}_{d,T}(\boldsymbol{\theta}_0)^{-1}\Big[(\boldsymbol{\tau} - \boldsymbol{\tau}_0) \odot \boldsymbol{d}_t(\boldsymbol{\theta}_0)\Big] \ln t\Big\|^2 = o(1),$

(iv) $\sup_{\gamma \in \mathcal{N}_{\delta,T}(\gamma_0)} \frac{1}{T} (\ln T)^k \sum_{t=1}^{T} \Big\|\boldsymbol{D}_{d,T}(\boldsymbol{\theta}_0)^{-1}\Big[(\boldsymbol{\tau} - \boldsymbol{\tau}_0) \odot (\boldsymbol{d}_t(\boldsymbol{\theta}) - \boldsymbol{d}_t(\boldsymbol{\theta}_0))\Big] \ln t\Big\|^2 = o(1),$

(v) $\sup_{\gamma \in \mathcal{N}_{\delta,T}(\gamma_0)} \sum_{t=1}^{T} \big|f(\boldsymbol{w}_t, \gamma) - f(\boldsymbol{w}_t, \gamma_0)\big|^2 = O_p((\ln T)^4),$

(vi) $\sup_{\gamma \in \mathcal{N}_{\delta,T}(\gamma_0)} \frac{1}{T^2} (\ln T)^k \sum_{t=1}^{T} \big\|\boldsymbol{D}_{d,T}(\boldsymbol{\theta}_0)^{-2} \operatorname{diag}[\boldsymbol{d}_t(\boldsymbol{\theta})]\big\|^2 = o(1),$

(vii) $\sup_{\gamma \in \mathcal{N}_{\delta,T}(\gamma_0)} \frac{1}{T} (\ln T) \big\|\sum_{t=1}^{T} \boldsymbol{D}_{d,T}(\boldsymbol{\theta}_0)^{-2} \operatorname{diag}[\boldsymbol{d}_t(\boldsymbol{\theta})] u_t (\ln t)^k\big\| = o_p(1).$

Proof When needed, let $i \in \{1, 2, \ldots, d\}$ be arbitrary. The i^{th} component of $\boldsymbol{\theta}_0$ and $\boldsymbol{\tau}_0$ are written θ_{0i} and τ_{0i}, respectively. ***(i)*** The i^{th} component of $\boldsymbol{D}_{d,T}(\boldsymbol{\theta}_0)^{-1}(\boldsymbol{d}_t(\boldsymbol{\theta}) - \boldsymbol{d}_t(\boldsymbol{\theta}_0))$ equals $T^{-\theta_{0i}}(t^{\theta_i} - t^{\theta_{0i}})$. By the mean-value theorem, we have

$$
\begin{aligned}
\frac{1}{T} (\ln T)^k \sum_{t=1}^{T} T^{-2\theta_{0i}} \big|t^{\theta_i} - t^{\theta_{0i}}\big|^2 &= \frac{1}{T} (\ln T)^k \sum_{t=1}^{T} \Big(\frac{t}{T}\Big)^{2\theta_{0i}} \big|t^{\theta_i - \theta_{0i}} - 1\big|^2 \\
&\leq \frac{1}{T} (\ln T)^k \sum_{t=1}^{T} \Big(\frac{t}{T}\Big)^{2\theta_{0i}} (\ln t)^2 \, |\theta_i - \theta_{0i}|^2 \leq \frac{1}{T} (\ln T)^{k+2} \sum_{t=1}^{T} \Big(\frac{t}{T}\Big)^{2\theta_{0i}} |\theta_i - \theta_{0i}|^2 \qquad \text{(A.26)} \\
&\leq (\ln T)^{k+2} \Bigg(\sup_{\theta_L \leq \theta \leq \theta_U} \frac{1}{T} \sum_{t=1}^{T} \Big(\frac{t}{T}\Big)^{2\theta}\Bigg) |\theta_i - \theta_{0i}|^2 .
\end{aligned}
$$

The supremum in the RHS of (A.26) is bounded in view of Lemma 1(a) in the supplement. Moreover, if $\theta_i \in \mathcal{N}_{\delta,T}(\gamma_0)$, then $T^{\theta_{0i}}\,|\theta_i - \theta_{0i}| < \delta T^{-1/2}\ln T$ or equivalently $|\theta_i - \theta_{0i}|^2 < T^{-(1+2\theta_{0i})}\ln T$. We conclude that

$$\frac{1}{T}(\ln T)^k \sum_{t=1}^{T} T^{-2\theta_{0i}} \left|t^{\theta_i} - t^{\theta_{0i}}\right| \le C(\ln T)^{k+3} T^{-(1+2\theta_{0i})} \tag{A.27}$$

which converges to zero because $1 + 2\theta_{0i} \ge 1 + 2\theta_L > 0$. The result follows since i was arbitrary. ***(ii)*** We again look at the i^{th} component of $\boldsymbol{D}_{d,T}(\boldsymbol{\theta}_0)^{-1}\Big[\boldsymbol{\tau}_0 \odot \big(\boldsymbol{d}_t(\boldsymbol{\theta}) - \boldsymbol{d}_t(\boldsymbol{\theta}_0)\big)\Big]\ln\frac{t}{T}$ and find

$$\frac{1}{T}(\ln T)^k \sum_{t=1}^{T} T^{-2\theta_{0i}}\tau_{0i}^2 \left|t^{\theta_i} - t^{\theta_{0i}}\right|^2 \left(\ln\frac{t}{T}\right)^2 \le (\ln T)^{k+2}\tau_{0i}^2 \left(\sup_{\theta_L\le\theta\le\theta_U} \frac{1}{T}\sum_{t=1}^{T}\left(\frac{t}{T}\right)^{2\theta}\left(\ln\frac{t}{T}\right)^2\right)|\theta_i - \theta_{0i}|^2.$$

taking steps identical to those taken in (A.26). The supremum in the RHS is bounded because $\sup_{\theta_L\le\theta\le\theta_U]}\left|\frac{1}{T}\sum_{t=1}^{T}\left(\frac{t}{T}\right)^{2\theta}\left(\ln\frac{t}{T}\right)^2 - \int_0^1 r^{2\theta}(\ln r)^2 dr\right| \to 0$ (a consequence of Lemma 1(iv)) and since $\int r^{2\theta_{0i}}(\ln r)^2 dr$ is finite for all $\theta \in [\theta_L, \theta_U]$. The proof is easily completed after recalling that $|\theta_i - \theta_{0i}|^2 < T^{-(1+2\theta_{0i})}\ln T$ whenever $\theta_i \in \mathcal{N}_{\delta,T}(\gamma_0)$. ***(iii)*** The contribution of the i^{th} component of $\boldsymbol{D}_{d,T}(\boldsymbol{\theta}_0)^{-1}[(\boldsymbol{\tau} - \boldsymbol{\tau}_0)\odot \boldsymbol{d}_t(\boldsymbol{\theta}_0)]\ln t$ to the sum $\frac{1}{T}(\ln T)^k\sum_{t=1}^{T}\left\|\boldsymbol{D}_{d,T}(\boldsymbol{\theta}_0)^{-1}[(\boldsymbol{\tau} - \boldsymbol{\tau}_0)\odot \boldsymbol{d}_t(\boldsymbol{\theta}_0)]\ln t\right\|^2$ is

$$\frac{1}{T}(\ln T)^k \sum_{t=1}^{T} T^{-2\theta_{0i}}(\tau_i - \tau_{0i})^2 t^{2\theta_{0i}}(\ln t)^2 \le (\ln T)^{k+2}\left(\sup_{\theta_L\le\theta\le\theta_U}\frac{1}{T}\sum_{t=1}^{T}\left(\frac{t}{T}\right)^{2\theta}\right)(\tau_i - \tau_{0i})^2. \tag{A.28}$$

The supremum is bounded so it remains to say something about $(\tau_i - \tau_{0i})^2$. By the triangle inequality and the properties of norms (namely $\|\operatorname{diag}[\boldsymbol{a}]\| \le \|\boldsymbol{a}\|$), we have

$$\begin{aligned}
&\left\|\boldsymbol{D}_{d,T}(\boldsymbol{\theta}_0)(\boldsymbol{\tau} - \boldsymbol{\tau}_0)\right\| \\
&\qquad \le \left\|\boldsymbol{D}_{d,T}(\boldsymbol{\theta}_0)\Big[\boldsymbol{\tau} - \boldsymbol{\tau}_0 + [\boldsymbol{\tau}_0 \odot (\boldsymbol{\theta} - \boldsymbol{\theta}_0)]\ln T\Big]\right\| + \left\|\boldsymbol{D}_{d,T}(\boldsymbol{\theta}_0)\operatorname{diag}[\boldsymbol{\tau}_0](\boldsymbol{\theta} - \boldsymbol{\theta}_0)\right\|\ln T \\
&\qquad \le \left\|\boldsymbol{D}_{d,T}(\boldsymbol{\theta}_0)\Big[\boldsymbol{\tau} - \boldsymbol{\tau}_0 + [\boldsymbol{\tau}_0 \odot (\boldsymbol{\theta} - \boldsymbol{\theta}_0)]\ln T\Big]\right\| + \|\boldsymbol{\tau}_0\|\left\|\boldsymbol{D}_{d,T}(\boldsymbol{\theta}_0)(\boldsymbol{\theta} - \boldsymbol{\theta}_0)\right\|\ln T \\
&\qquad \le \delta(1 + \|\boldsymbol{\tau}_0\|)T^{-1/2}(\ln T)^2,
\end{aligned} \tag{A.29}$$

for all $\boldsymbol{\tau} \in \mathcal{N}_{\delta,T}(\gamma_0)$. (A.29) implies that $(\tau_i - \tau_{0i})^2 \le \delta(1 + \|\boldsymbol{\tau}_0\|)(\ln T)^4 T^{-(1+2\theta_{0i})}$ which goes to zero as $T \to \infty$. Now combine this finding with the RHS of (A.28) to establish the result. ***(iv)*** Use similar arguments as used in the proofs of (i) and (iii). ***(v)*** By definition of $f(\boldsymbol{w}_t, \gamma)$ (see Lemma 3), it follows that

$$\begin{aligned}
f(\boldsymbol{w}_t,\gamma) - f(\boldsymbol{w}_t,\gamma_0) &= \boldsymbol{d}_t(\boldsymbol{\theta})'\boldsymbol{\tau} + \boldsymbol{s}_t'\boldsymbol{\phi} - [\boldsymbol{d}_t(\boldsymbol{\theta}_0)'\boldsymbol{\tau}_0 + \boldsymbol{s}_t'\boldsymbol{\phi}_0] \\
&= [\boldsymbol{d}_t(\theta) - \boldsymbol{d}_t(\theta_0)]'(\boldsymbol{\tau} - \boldsymbol{\tau}_0) + \boldsymbol{d}_t(\boldsymbol{\theta}_0)'(\boldsymbol{\tau} - \boldsymbol{\tau}_0) + [\boldsymbol{d}_t(\theta) - \boldsymbol{d}_t(\theta_0)]'\boldsymbol{\tau}_0 + \boldsymbol{s}_t'(\boldsymbol{\phi} - \boldsymbol{\phi}_0).
\end{aligned}$$

and by the c_r-inequality that

$$\begin{aligned}
\sum_{t=1}^{T}\left|f(\boldsymbol{w}_t,\gamma) - f(\boldsymbol{w}_t,\gamma_0)\right|^2 \le C\Bigg\{&\sum_{t=1}^{T}\left|(\boldsymbol{d}_t(\boldsymbol{\theta}) - \boldsymbol{d}_t(\boldsymbol{\theta}_0))'(\boldsymbol{\tau} - \boldsymbol{\tau}_0)\right|^2 + \sum_{t=1}^{T}|\boldsymbol{d}_t(\boldsymbol{\theta}_0)'(\boldsymbol{\tau} - \boldsymbol{\tau}_0)|^2 \\
&+ \sum_{t=1}^{T}\left|(\boldsymbol{d}_t(\boldsymbol{\theta}) - \boldsymbol{d}_t(\boldsymbol{\theta}_0))'\boldsymbol{\tau}_0\right|^2 + \sum_{t=1}^{T}\left|\boldsymbol{s}_t'(\boldsymbol{\phi} - \boldsymbol{\phi}_0)\right|^2\Bigg\} =: C\{IVa + IVb + IVc + IVd\}.
\end{aligned} \tag{A.30}$$

It remains to bound the terms IVa-IVd uniformly over $\mathcal{N}_{\delta,T}(\gamma_0)$. We repeatedly rely on $|\boldsymbol{a}'\boldsymbol{b}|^2 \leq \|\boldsymbol{a}\|^2\|\boldsymbol{b}\|^2$ (Cauchy-Schwartz). We have

$$
\begin{aligned}
IVa &= \sum_{t=1}^{T} \left|[\boldsymbol{D}_{d,T}(\boldsymbol{\theta}_0)^{-1}(\boldsymbol{d}_t(\boldsymbol{\theta}) - \boldsymbol{d}_t(\boldsymbol{\theta}_0))]'[\boldsymbol{D}_{d,T}(\boldsymbol{\theta}_0)(\boldsymbol{\tau} - \boldsymbol{\tau}_0)]\right|^2 \\
&\leq T \left\|\boldsymbol{D}_{d,T}(\boldsymbol{\theta}_0)(\boldsymbol{\tau} - \boldsymbol{\tau}_0)\right\|^2 \frac{1}{T}\sum_{t=1}^{T} \left\|\boldsymbol{D}_{d,T}(\boldsymbol{\theta}_0)^{-1}(\boldsymbol{d}_t(\boldsymbol{\theta}) - \boldsymbol{d}_t(\boldsymbol{\theta}_0))\right\|^2 \to 0
\end{aligned}
\tag{A.31}
$$

on $\mathcal{N}_{\delta,T}(\gamma_0)$ as $T \to \infty$ by (A.29) and Lemma 4(a). The bound on IVb is derived similarly, that is

$$
\begin{aligned}
IVb &\leq T \left\|\boldsymbol{D}_{d,T}(\boldsymbol{\theta}_0)(\boldsymbol{\tau} - \boldsymbol{\tau}_0)\right\|^2 \frac{1}{T}\sum_{t=1}^{T} \left\|\boldsymbol{D}_{d,T}(\boldsymbol{\theta}_0)^{-1}\boldsymbol{d}_t(\boldsymbol{\theta}_0)\right\|^2 \\
&= T \left\|\boldsymbol{D}_{d,T}(\boldsymbol{\theta}_0)(\boldsymbol{\tau} - \boldsymbol{\tau}_0)\right\|^2 \frac{1}{T}\sum_{t=1}^{T}\sum_{i=1}^{d} \left(\frac{t}{T}\right)^{2\theta_{0i}} \leq dT \left\|\boldsymbol{D}_{d,T}(\boldsymbol{\theta}_0)(\boldsymbol{\tau} - \boldsymbol{\tau}_0)\right\|^2 \frac{1}{T}\sum_{t=1}^{T}\left(\frac{t}{T}\right)^{2\theta_L},
\end{aligned}
\tag{A.32}
$$

which is $O((\ln T)^4)$ uniformly over $\mathcal{N}_{\delta,T}(\gamma_0)$ because $\left\|\boldsymbol{D}_{d,T}(\boldsymbol{\theta}_0)(\boldsymbol{\tau} - \boldsymbol{\tau}_0)\right\|^2 = O(T^{-1}(\ln T)^4)$ (using (A.29)) and $\frac{1}{T}\sum_{t=1}^{T}\left(\frac{t}{T}\right)^{2\theta_L}$ is bounded (see Lemma 1(i)). For the third term, we establish

$$
\begin{aligned}
IVc &\leq \|\boldsymbol{\tau}_0\|^2 \sum_{t=1}^{T}\sum_{i=1}^{d} \left|t^{\theta_i} - t^{\theta_{0i}}\right|^2 = \|\boldsymbol{\tau}_0\|^2 \sum_{i=1}^{d}\sum_{t=1}^{T} t^{2\theta_{0i}}\left|t^{\theta_i - \theta_{0i}} - 1\right|^2 \leq \|\boldsymbol{\tau}_0\|^2 \sum_{i=1}^{d}\sum_{t=1}^{T} t^{2\theta_{0i}}(\ln t)^2(\theta_i - \theta_{0i})^2 \\
&\leq \|\boldsymbol{\tau}_0\|^2(\ln T)^2 T \sum_{i=1}^{d} T^{2\theta_{0i}}(\theta_i - \theta_{0i})^2 \left[\frac{1}{T}\sum_{t=1}^{T}\left(\frac{t}{T}\right)^{2\theta_L}\right] \\
&= \|\boldsymbol{\tau}_0\|^2(\ln T)^2 T \left\|\boldsymbol{D}_{d,T}(\boldsymbol{\theta}_0)(\boldsymbol{\theta} - \boldsymbol{\theta}_0)\right\|^2 \left[\frac{1}{T}\sum_{t=1}^{T}\left(\frac{t}{T}\right)^{2\theta_L}\right] = O((\ln T)^4)
\end{aligned}
$$

using the mean-value theorem, Lemma 1(i), and the fact that $\left\|\boldsymbol{D}_{d,T}(\boldsymbol{\theta}_0)(\boldsymbol{\theta} - \boldsymbol{\theta}_0)\right\|^2 = O(T^{-1}(\ln T)^2)$ on $\mathcal{N}_{\delta,T}(\gamma_0)$. Finally, the bound on IVd. We have

$$
IVd \leq T \left\|\boldsymbol{D}_{s,T}(\boldsymbol{\phi} - \boldsymbol{\phi}_0)\right\|^2 \frac{1}{T}\sum_{t=1}^{T} \|\boldsymbol{D}_{s,T}^{-1}\boldsymbol{s}_t\|^2 \leq T \left\|\boldsymbol{D}_{s,T}(\boldsymbol{\phi} - \boldsymbol{\phi}_0)\right\|^2 \sum_{i=1}^{m}\sum_{j=1}^{p_i} \frac{1}{T}\sum_{t=1}^{T}\left(\frac{x_{it}}{\sqrt{T}}\right)^{2j},
$$

which is $O_p((\ln T)^2)$ since $\left\|\boldsymbol{D}_{s,T}(\boldsymbol{\phi} - \boldsymbol{\phi}_0)\right\|^2 \leq \delta T^{-1}(\ln T)^2$ on $\mathcal{N}_{\delta,T}(\gamma_0)$ and since Assumption 2 guarantees that all terms of the form $\frac{1}{T}\sum_{t=1}^{T}(x_{it}/\sqrt{T})^{2j}$ converge to integrals of Brownian motions.
(vi) It follows by $\|\operatorname{diag}[\boldsymbol{a}]\| \leq \|\boldsymbol{a}\|$, $\|\boldsymbol{AB}\| \leq \|\boldsymbol{A}\|\,\|\boldsymbol{B}\|$, the c_r-inequality, and the triangle inequality that

$$
\begin{aligned}
\frac{1}{T^2}(\ln T)^k \sum_{t=1}^{T} \left\|\boldsymbol{D}_{d,T}(\boldsymbol{\theta}_0)^{-2}\operatorname{diag}[\boldsymbol{d}_t(\boldsymbol{\theta})]\right\|^2 &\leq \frac{1}{T^2}(\ln T)^k \sum_{t=1}^{T} \left\|\boldsymbol{D}_{d,T}(\boldsymbol{\theta}_0)^{-2}\boldsymbol{d}_t(\boldsymbol{\theta})\right\|^2 \\
&\leq C\frac{1}{T}(\ln T)^k \left\|\boldsymbol{D}_{d,T}(\boldsymbol{\theta}_0)^{-1}\right\| \frac{1}{T}\sum_{t=1}^{T} \left\|\boldsymbol{D}_{d,T}(\boldsymbol{\theta}_0)^{-1}(\boldsymbol{d}_t(\boldsymbol{\theta}) - \boldsymbol{d}_t(\boldsymbol{\theta}_0))\right\|^2 \\
&\quad + C\frac{1}{T}(\ln T)^k \left\|\boldsymbol{D}_{d,T}(\boldsymbol{\theta}_0)^{-1}\right\| \frac{1}{T}\sum_{t=1}^{T} \left\|\boldsymbol{D}_{d,T}(\boldsymbol{\theta}_0)^{-1}\boldsymbol{d}_t(\boldsymbol{\theta}_0)\right\|^2
\end{aligned}
\tag{A.33}
$$

Both terms are negligible since $\frac{1}{T}(\ln T)^k \left\|\boldsymbol{D}_{d,T}(\boldsymbol{\theta}_0)^{-1}\right\| \leq (\ln T)^k T^{-(1+\theta_L)} \to 0$, $\frac{1}{T}\sum_{t=1}^T \left\|\boldsymbol{D}_{d,T}(\boldsymbol{\theta}_0)^{-1}(\boldsymbol{d}_t(\boldsymbol{\theta}) - \boldsymbol{d}_t(\boldsymbol{\theta}_0))\right\|^2$ is $o(1)$ on $\mathcal{N}_{\delta,T}(\gamma_0)$ (Lemma 4(i)), and $\frac{1}{T}\sum_{t=1}^T \left\|\boldsymbol{D}_{d,T}(\boldsymbol{\theta}_0)^{-1}\boldsymbol{d}_t(\boldsymbol{\theta}_0)\right\|^2$ is bounded in view of Lemma 1(i). ***(vii)*** Use similar steps as in (A.33) to obtain

$$T^{-1}(\ln T)\left\|\sum_{t=1}^T \boldsymbol{D}_{d,T}(\boldsymbol{\theta}_0)^{-2}\,\mathrm{diag}[\boldsymbol{d}_t(\boldsymbol{\theta})]u_t(\ln t)^k\right\| \leq (\ln T)T^{-(\theta_L+\frac{1}{2})}\times$$
$$\Bigg\{\underbrace{\left\|T^{-1/2}\sum_{t=1}^T \boldsymbol{D}_{d,T}(\boldsymbol{\theta}_0)^{-1}\boldsymbol{d}_t(\boldsymbol{\theta}_0)u_t(\ln t)^k\right\|}_{V(a)} + \underbrace{\left\|T^{-1/2}\sum_{t=1}^T \boldsymbol{D}_{d,T}(\boldsymbol{\theta}_0)^{-1}(\boldsymbol{d}_t(\boldsymbol{\theta}) - \boldsymbol{d}_t(\boldsymbol{\theta}_0))u_t(\ln t)^k\right\|}_{V(b)}\Bigg\}.$$

Note that the i^{th} component of the vector in $V(a)$ equals $T^{-1/2}\sum_{t=1}^T \left(\frac{t}{T}\right)^{\theta_{0i}} u_t(\ln t)^k$. Lemma 1(iii) and Chebyshev's inequality imply that $V(a)$ is $O_p((\ln T)^k)$. Furthermore, the mean-value theorem implies

$$\begin{aligned}
\left\|V(b)\right\| &\leq \left\|[\boldsymbol{D}_{d,T}(\boldsymbol{\theta}_0)(\boldsymbol{\theta} - \boldsymbol{\theta}_0)] \odot \left[T^{-1/2}\sum_{t=1}^T \boldsymbol{D}_{d,T}(\boldsymbol{\theta}_0)^{-1}\boldsymbol{d}_t(\tilde{\boldsymbol{\theta}})u_t(\ln t)^{k+1}\right]\right\| \\
&\leq \left\|\boldsymbol{D}_{d,T}(\boldsymbol{\theta}_0)(\boldsymbol{\theta} - \boldsymbol{\theta}_0)\right\| \left\|T^{-1/2}\sum_{t=1}^T \boldsymbol{D}_{d,T}(\boldsymbol{\theta}_0)^{-1}\boldsymbol{d}_t(\tilde{\boldsymbol{\theta}})u_t(\ln t)^{k+1}\right\| \\
&\leq \left\|\boldsymbol{D}_{d,T}(\boldsymbol{\theta}_0)(\boldsymbol{\theta} - \boldsymbol{\theta}_0)\right\| \left\|\boldsymbol{D}_{d,T}(\boldsymbol{\theta}_0)\boldsymbol{D}_{d,T}(\tilde{\boldsymbol{\theta}})\right\| \left\|T^{-1/2}\sum_{t=1}^T \boldsymbol{D}_{d,T}(\boldsymbol{\theta}_0)^{-1}\boldsymbol{d}_t(\tilde{\boldsymbol{\theta}})u_t(\ln t)^{k+1}\right\| \\
&\leq \left(\delta(\ln T)T^{-1/2}\right) T^{\|\boldsymbol{\theta}-\boldsymbol{\theta}_0\|_\infty}\, O_p((\ln T)^{k+2})
\end{aligned} \tag{A.34}$$

using Lemma 1(iii) and $\boldsymbol{\theta} \in \mathcal{N}_{\delta,T}(\gamma_0)$. Having established the stochastic orders of $V(a)$ and $V(b)$ it is straightforward to verify the claim in (vii). ■

B Proof of Main Theorems

C Proof of Theorem 1

C.1 Theorem 1

This section is devoted to the proof of Theorems 1 - 4. Before continuing, we recall the criterion function being defined as $Q_T(\gamma) = \frac{1}{2}\sum_{t=1}^T (y_t - f(\boldsymbol{w}_t, \gamma))^2$ with $f(\boldsymbol{w}_t, \gamma) = \boldsymbol{d}_t(\boldsymbol{\theta})'\boldsymbol{\tau} + \boldsymbol{s}_t'\boldsymbol{\phi}$ and $\boldsymbol{w}_t = [t, \boldsymbol{x}_t]'$. From theorem 3.1 in Chan and Wang (2015), we have $\boldsymbol{G}_{\gamma_0,T}(\widehat{\gamma}_T - \gamma_0) = \boldsymbol{M}_T^{-1}\boldsymbol{z}_T + o_p(1)$, if the following five conditions are fulfilled:[17]

(i) $\left\|\boldsymbol{G}_{\gamma_0,T}^{-1}\right\| \to 0$ as $T \to \infty$,

(ii) $\sup_{\gamma:\,\|\boldsymbol{G}_{\gamma_0,T}(\gamma-\gamma_0)\|\leq k_T} \left\|\boldsymbol{G}_{\gamma_0,T}'^{-1}\sum_{t=1}^T \left[\dot{f}(\boldsymbol{w}_t, \gamma)\dot{f}(\boldsymbol{w}_t, \gamma)' - \dot{f}(\boldsymbol{w}_t, \gamma_0)\dot{f}(\boldsymbol{w}_t, \gamma_0)'\right]\boldsymbol{G}_{\gamma_0,T}^{-1}\right\| = o_p(1)$,

(iii) $\sup_{\gamma:\,\|\boldsymbol{G}_{\gamma_0,T}(\gamma-\gamma_0)\|\leq k_T} \left\|\boldsymbol{G}_{\gamma_0,T}'^{-1}\sum_{t=1}^T \ddot{f}(\boldsymbol{w}_t, \gamma)\left[f(\boldsymbol{w}_t, \gamma) - f(\boldsymbol{w}_t, \gamma_0)\right]\boldsymbol{G}_{\gamma_0,T}^{-1}\right\| = o_p(1)$,

[17] The original result in theorem 3.1 of Chan and Wang (2015) does not explicitly allow for deterministic trends and a scaling matrix $\boldsymbol{G}_{\gamma_0,T}$ that is parameter dependent. However, all steps in the proof remain valid after allowing for these features.

(iv) $\sup_{\gamma:\,\|G_{\gamma_0,T}(\gamma-\gamma_0)\|\le k_T} \left\|G'^{-1}_{\gamma_0,T}\sum_{t=1}^T \ddot{f}(w_t,\gamma)u_t G^{-1}_{\gamma_0,T}\right\| = o_p(1)$;

(v) for any $\alpha_i = [\alpha_{i1},\dots,\alpha_{i,2d+p}]' \in \mathbb{R}^{2d+p}$, $i = 1,2,3$,

$$[\alpha_1' M_T \alpha_2, \alpha_3' z_T] \Rightarrow [\alpha_1' M \alpha_2, \alpha_3' z],$$

where $M > 0$ a.s., $\mathbb{P}(z < \infty) = 1$ and

$$M_T = G'^{-1}_{\gamma_0,T}\left[\sum_{t=1}^T \dot{f}(w_t,\gamma_0)\dot{f}(w_t,\gamma_0)'\right]G^{-1}_{\gamma_0,T}, \qquad z_T = G'^{-1}_{\gamma_0,T}\left[\sum_{t=1}^T \dot{f}(w_t,\gamma_0)u_t\right]. \quad \text{(C.1)}$$

We make two remarks before verifying these conditions. First, we set $k_T = \delta \ln T$ and we will verify conditions (ii)-(iv) while replacing the supremum over the set

$$\widetilde{\mathcal{N}}_{\delta,T}(\gamma_0) = \{\gamma \in \Gamma : \|G_{\gamma_0,T}(\gamma-\gamma_0)\| \le \delta \ln T\}$$

by a supremum over the set $\mathcal{N}_{\delta,T}(\gamma_0)$ given in (A.25).

$$\begin{aligned}\mathcal{N}_{\delta,T}(\gamma_0) = \Big\{\gamma \in \Gamma :\ &\left\|D_{d,T}(\theta_0)(\theta-\theta_0)\right\| < \delta T^{-1/2}\ln T, \left\|D_{s,T}(\phi-\phi_0)\right\| < \delta T^{-1/2}\ln T,\\ &\left\|D_{d,T}(\theta_0)\big((\tau-\tau_0) + [\tau_0 \odot (\theta-\theta_0)]\ln T\big)\right\| < \delta T^{-1/2}\ln T\Big\}.\end{aligned} \quad \text{(C.2)}$$

Since $\widetilde{\mathcal{N}}_{\delta,T}(\gamma_0) \subset \mathcal{N}_{\delta,T}(\gamma_0)$, this replacement is innocuous. Second, note that $\|a \odot b\| \le \|a \odot b\|_1 \le \|a\|\,\|b\|$ holds for comformable vectors a and b.

Proof of Theorem 1 ***(i)*** From $\|L_{\tau_0,T}\| \le \|L_{\tau_0,T}\|_{\mathcal{F}} = \left(2d+p+2\|\tau_0\|^2(\ln T)^2\right)^{1/2} \le C\ln T$, we have

$$\left\|G^{-1}_{\gamma_0,T}\right\| \le \left\|L_{\tau_0,T}\right\|\left\|D^{-1}_{\theta_0,T}\right\| \le C(\ln T)T^{-1/2}(T^{-\theta_l} + T^{-1/2}) \to 0,$$

as $T \to \infty$. ***(ii)*** Use $\sum_t a_t a_t' - \sum_t b_t b_t' = \sum_t (a_t - b_t)(a_t - b_t)' + \sum_t (a_t - b_t)b_t' + \sum_t b_t(a_t - b_t)'$ to obtain the bound

$$\begin{aligned}&\left\|G'^{-1}_{\gamma_0,T}\sum_{t=1}^T\left[\dot{f}(w_t,\gamma)\dot{f}(w_t,\gamma)' - \dot{f}(w_t,\gamma_0)\dot{f}(w_t,\gamma_0)'\right]G^{-1}_{\gamma_0,T}\right\|\\
&\le \left\|D^{-1}_{\theta_0,T}\left[\sum_{t=1}^T L_{\tau_0,T}\left(\dot{f}(w_t,\gamma)-\dot{f}(w_t,\gamma_0)\right)\left(\dot{f}(w_t,\gamma)-\dot{f}(w_t,\gamma_0)\right)'L'_{\tau_0,T}\right]D^{-1}_{\theta_0,T}\right\|\\
&\quad + 2\left\|D^{-1}_{\theta_0,T}\left[\sum_{t=1}^T L_{\tau_0,T}\left(\dot{f}(w_t,\gamma)-\dot{f}(w_t,\gamma_0)\right)\dot{f}(w_t,\gamma_0)'L'_{\tau_0,T}\right]D^{-1}_{\theta_0,T}\right\| \qquad \text{(C.3)}\\
&\le \sum_{t=1}^T\left\|D^{-1}_{\theta_0,T}L_{\tau_0,T}\left(\dot{f}(w_t,\gamma)-\dot{f}(w_t,\gamma_0)\right)\right\|^2\\
&\quad + 2\sqrt{(\ln T)^2\sum_{t=1}^T\left\|D^{-1}_{\theta_0,T}L_{\tau_0,T}\left(\dot{f}(w_t,\gamma)-\dot{f}(w_t,\gamma_0)\right)\right\|^2}\sqrt{(\ln T)^{-2}\sum_{t=1}^T\left\|D^{-1}_{\theta_0,T}L_{\tau_0,T}\dot{f}(w_t,\gamma_0)\right\|^2}.\end{aligned}$$

For any $k \geq 0$, the c_r-inequality and (A.22) yield a further upper bound as

$$
\begin{aligned}
&(\ln T)^k \sum_{t=1}^{T} \left\| \boldsymbol{D}_{\boldsymbol{\theta}_0,T}^{-1} \boldsymbol{L}_{\boldsymbol{\tau}_0,T} \left(\dot{f}(\boldsymbol{w}_t,\gamma) - \dot{f}(\boldsymbol{w}_t,\gamma_0) \right) \right\|^2 \\
&\leq C(\ln T)^k \left\{ \frac{1}{T} \sum_{t=1}^{T} \left\| \boldsymbol{D}_{d,T}(\boldsymbol{\theta}_0)^{-1} \Big((\boldsymbol{\tau} - \boldsymbol{\tau}_0) \odot (\boldsymbol{d}_t(\boldsymbol{\theta}) - \boldsymbol{d}_t(\boldsymbol{\theta}_0)) \Big) \ln t \right\|^2 \right. \\
&+ \frac{1}{T} \sum_{t=1}^{T} \left\| \boldsymbol{D}_{d,T}(\boldsymbol{\theta}_0)^{-1} (\boldsymbol{d}_t(\boldsymbol{\theta}) - \boldsymbol{d}_t(\boldsymbol{\theta}_0)) \right\|^2 + \frac{1}{T} \sum_{t=1}^{T} \left\| \boldsymbol{D}_{d,T}(\boldsymbol{\theta}_0)^{-1} \Big(\boldsymbol{\tau}_0 \odot (\boldsymbol{d}_t(\boldsymbol{\theta}) - \boldsymbol{d}_t(\boldsymbol{\theta}_0)) \Big) \ln \frac{t}{T} \right\|^2 \\
&\left. + \frac{1}{T} \sum_{t=1}^{T} \left\| \boldsymbol{D}_{d,T}(\boldsymbol{\theta}_0)^{-1} \Big((\boldsymbol{\tau} - \boldsymbol{\tau}_0) \odot \boldsymbol{d}_t(\boldsymbol{\theta}_0) \Big) \ln t \right\|^2 \right\}.
\end{aligned}
\tag{C.4}
$$

It follows from properties (i)-(iv) of Lemma 4 that each term in the RHS of (C.4) is $o(1)$ uniformly over $\mathcal{N}_{\delta,T}(\gamma_0)$. Moreover, since $\left(\ln \frac{t}{T}\right)^2 \leq 2(\ln T)^2$ for every $t = 1, 2, \ldots, T$, we find that

$$
\begin{aligned}
(\ln T)^{-2} \sum_{t=1}^{T} \left\| \boldsymbol{D}_{\boldsymbol{\theta}_0,T}^{-1} \boldsymbol{L}_{\boldsymbol{\tau}_0,T} \dot{f}(\boldsymbol{w}_t,\gamma_0) \right\|^2 &= (\ln T)^{-2} \frac{1}{T} \sum_{t=1}^{T} \left\| \boldsymbol{D}_{d,T}(\boldsymbol{\theta}_0)^{-1} \Big(\boldsymbol{\tau}_0 \odot \boldsymbol{d}_t(\boldsymbol{\theta}_0) \Big) \ln \frac{t}{T} \right\|^2 \\
&+ (\ln T)^{-2} \frac{1}{T} \sum_{t=1}^{T} \left\| \boldsymbol{D}_{d,T}(\boldsymbol{\theta}_0)^{-1} \boldsymbol{d}_t(\boldsymbol{\theta}_0) \right\|^2 + (\ln T)^{-2} \frac{1}{T} \sum_{t=1}^{T} \left\| \boldsymbol{D}_{s,T}^{-1} \boldsymbol{s}_t \right\|^2
\end{aligned}
\tag{C.5}
$$

is $O_p(1)$ in view of Lemmas 1(i) and 2(iii). The combination of (C.3), (C.4), and (C.5) leads to the desired result. ***(iii)*** The Cauchy-Schwarz inequality implies

$$
\begin{aligned}
&\left\| \boldsymbol{G}_{\gamma_0,T}^{\prime -1} \sum_{t=1}^{T} \ddot{f}(\boldsymbol{w}_t,\gamma) \left[f(\boldsymbol{w}_t,\gamma) - f(\boldsymbol{w}_t,\gamma_0) \right] \boldsymbol{G}_{\gamma_0,T}^{-1} \right\| \\
&\qquad \leq \sqrt{(\ln T)^4 \sum_{t=1}^{T} \left\| \boldsymbol{G}_{\gamma_0,T}^{\prime -1} \ddot{f}(\boldsymbol{w}_t,\gamma) \boldsymbol{G}_{\gamma_0,T}^{-1} \right\|^2} \sqrt{(\ln T)^{-4} \sum_{t=1}^{T} \left| f(\boldsymbol{w}_t,\gamma) - f(\boldsymbol{w}_t,\gamma_0) \right|^2}.
\end{aligned}
\tag{C.6}
$$

Using the identity in Lemma 3(v), we can bound the first term in the RHS of (C.6) as in

$$
\begin{aligned}
(\ln T)^4 \sum_{t=1}^{T} \left\| \boldsymbol{G}_{\gamma_0,T}^{\prime -1} \ddot{f}(\boldsymbol{w}_t,\gamma) \boldsymbol{G}_{\gamma_0,T}^{-1} \right\|^2 &\leq C(\ln T)^4 \sum_{t=1}^{T} \left\| \ddot{\boldsymbol{F}}_{11,t} \right\|^2 + C(\ln T)^4 \sum_{t=1}^{T} \left\| \ddot{\boldsymbol{F}}_{12,t} \right\|^2 \\
&\leq C(\ln T)^8 \frac{1}{T^2} \sum_{t=1}^{T} \left\| \boldsymbol{D}_{d,T}(\boldsymbol{\theta}_0)^{-2} \operatorname{diag}[\boldsymbol{d}_t(\boldsymbol{\theta})] \right\|^2,
\end{aligned}
\tag{C.7}
$$

which is $o(1)$ uniformly over $\mathcal{N}_{\delta,T}(\gamma_0)$ by Lemma 4(vi). Note that the second inequality in (C.7) makes use of the facts that: (1) all matrices in $\ddot{\boldsymbol{F}}_{11,t}$ and $\ddot{\boldsymbol{F}}_{12,t}$ are diagonal and therefore commute, and (2) the triangle inequality gives

$$
\|\operatorname{diag}[\boldsymbol{\tau}]\| \leq \|\boldsymbol{\tau}\| \leq \left\| \boldsymbol{D}_{d,T}(\boldsymbol{\theta}_0)^{-1} \right\| \left\| \boldsymbol{D}_{d,T}(\boldsymbol{\theta}_0)(\boldsymbol{\tau} - \boldsymbol{\tau}_0) \right\| + \|\boldsymbol{\tau}_0\| \quad \leq C(\ln T)^2 T^{-(\theta_l + 1/2)} + \|\boldsymbol{\tau}_0\| \leq C
$$

when $\gamma \in \mathcal{N}_{\delta,T}(\gamma_0)$ and T sufficiently large.[18] The first term in (C.6) is $o(1)$ and the second is $O(1)$ (see Lemma 4(v)) over $\mathcal{N}_{\delta,T}(\gamma_0)$. The claim follows. ***(iv)*** By Lemma 3(v), we have

$$
\left\| \boldsymbol{G}_{\gamma_0,T}^{\prime -1} \sum_{t=1}^{T} \ddot{f}(\boldsymbol{w}_t,\gamma) u_t \boldsymbol{G}_{\gamma_0,T}^{-1} \right\| \leq C \left(\left\| \sum_{t=1}^{T} \ddot{\boldsymbol{F}}_{11,t} u_t \right\| + \left\| \sum_{t=1}^{T} \ddot{\boldsymbol{F}}_{12,t} u_t \right\| \right)
\tag{C.8}
$$

[18]As an example, we consider $(\ln T)^4 \sum_{t=1}^{T} \left\| \ddot{\boldsymbol{F}}_{11,t} \right\|^2$ explicitly. Using the definition of $\ddot{\boldsymbol{F}}_{11,t}$ in Lemma 3(v) and the triangle inequality, we have $(\ln T)^4 \sum_{t=1}^{T} \left\| \ddot{\boldsymbol{F}}_{11,t} \right\|^2 \leq C \frac{(\ln T)^8}{T^2} \sum_{t=1}^{T} \left\| \boldsymbol{D}_{d,T}(\boldsymbol{\theta}_0)^{-1} \operatorname{diag}[\boldsymbol{\tau}] \operatorname{diag}[\boldsymbol{d}_t(\boldsymbol{\theta})] \boldsymbol{D}_{d,T}(\boldsymbol{\theta}_0)^{-1} \right\|^2 \leq C \frac{(\ln T)^8}{T^2} \|\operatorname{diag}[\boldsymbol{\tau}]\|^2 \sum_{t=1}^{T} \left\| \boldsymbol{D}_{d,T}(\boldsymbol{\theta}_0)^{-1} \operatorname{diag}[\boldsymbol{d}_t(\boldsymbol{\theta})] \boldsymbol{D}_{d,T}(\boldsymbol{\theta}_0)^{-1} \right\|^2$.

Using the definitions in Lemma 3(v) in a similar way as before, we find the inequalities

$$
\begin{aligned}
\left\|\sum_{t=1}^{T} \ddot{\boldsymbol{F}}_{11,t}u_t\right\| \leq \|\operatorname{diag}[\boldsymbol{\tau}]\| &\left\|T^{-1}\sum_{t=1}^{T}\boldsymbol{D}_{d,T}(\boldsymbol{\theta}_0)^{-2}\operatorname{diag}[\boldsymbol{d}_t(\boldsymbol{\theta})]u_t(\ln t)^2\right\| \\
&+ 2\|\operatorname{diag}[\boldsymbol{\tau}_0]\|(\ln T)\left\|T^{-1}\sum_{t=1}^{T}\boldsymbol{D}_{d,T}(\boldsymbol{\theta}_0)^{-2}\operatorname{diag}[\boldsymbol{d}_t(\boldsymbol{\theta})]u_t\ln t\right\|,
\end{aligned}
\tag{C.9}
$$

and $\left\|\sum_{t=1}^{T}\ddot{\boldsymbol{F}}_{12,t}u_t\right\| = \left\|T^{-1}\sum_{t=1}^{T}\boldsymbol{D}_{d,T}(\boldsymbol{\theta}_0)^{-2}\operatorname{diag}[\boldsymbol{d}_t(\boldsymbol{\theta})]u_t\ln t\right\|$. All relevant terms are $o_p(1)$ over $\mathcal{N}_{\delta,T}(\gamma_0)$ by Lemma 4(vii). ***(v)*** The convergence results in Lemma 2 applied to the definitions in Lemma 3(vi) provide $(\boldsymbol{M}_T, \boldsymbol{z}_T) \Rightarrow \left(\int_0^1 \boldsymbol{j}(r;\gamma_0)\boldsymbol{j}(r;\gamma_0)'dr,\ \int_0^1 \boldsymbol{j}(r;\gamma_0)dB_u(r) + \boldsymbol{\mathcal{B}}_{vu}\right)$ as $T\to\infty$. ■

C.2 Theorem 2

Proof of Theorem 2 Changing the summation indices, we can express the one-sided long-run covariance estimator as

$$
\widehat{\boldsymbol{\Delta}}_T(\widehat{\gamma}_T, b_T) = \sum_{i=0}^{T-1} k\left(\frac{i}{b_T}\right)\left(\frac{1}{T}\sum_{t=1}^{T-i}\boldsymbol{V}_{t+i}(\widehat{\gamma}_T)\boldsymbol{V}_t(\widehat{\gamma}_T)'\right),
\tag{C.10}
$$

where we explicitly indicate the dependence on the parameter estimator $\widehat{\gamma}_T$ and bandwidth b_T. If we define $\widehat{\boldsymbol{\Sigma}}_T(\widehat{\gamma}_T) = \frac{1}{T}\sum_{t=1}^{T}\boldsymbol{V}_t(\widehat{\gamma}_T)\boldsymbol{V}_t(\widehat{\gamma}_T)'$ and $\widehat{\boldsymbol{\Gamma}}_T(\widehat{\gamma}_T, b_T) = \sum_{i=1}^{T-1}k\left(\frac{i}{b_T}\right)\left(\frac{1}{T}\sum_{t=1}^{T-i}\boldsymbol{V}_t(\widehat{\gamma}_T)\boldsymbol{V}_t(\widehat{\gamma}_T)'\right)$, then $\widehat{\boldsymbol{\Delta}}_T(\widehat{\gamma}_T, b_T) = \widehat{\boldsymbol{\Sigma}}_T(\widehat{\gamma}_T) + \widehat{\boldsymbol{\Gamma}}_T(\widehat{\gamma}_T, b_T)$. We make two observations. First, the two-sided long-run covariance matrix estimator can be written as

$$
\widehat{\boldsymbol{\Omega}}_T(\widehat{\gamma}_T, b_T) = \widehat{\boldsymbol{\Sigma}}_T(\widehat{\gamma}_T) + \widehat{\boldsymbol{\Gamma}}_T(\widehat{\gamma}_T, b_T) + \widehat{\boldsymbol{\Gamma}}_T(\widehat{\gamma}_T, b_T)'.
\tag{C.11}
$$

It thus suffices to study the asymptotic behavior of $\widehat{\boldsymbol{\Sigma}}_T(\widehat{\gamma}_T)$ and $\widehat{\boldsymbol{\Gamma}}_T(\widehat{\gamma}_T, b_T)$ only. Second, the bottom right subblock of $\boldsymbol{V}_t(\gamma)\boldsymbol{V}_t(\gamma)'$ equals $\boldsymbol{v}_t\boldsymbol{v}_t'$ (no parameter estimation uncertainty here). The consistency results for this subblock are immediate from theorem 2 of Jansson (2002). We will therefore restrict our attention to $(1,1)^{th}$ elements of $\widehat{\boldsymbol{\Sigma}}_T(\widehat{\gamma}_T)$ and $\widehat{\boldsymbol{\Gamma}}_T(\widehat{\gamma}_T, b_T)$. That is, we will show

$$
\left[\widehat{\boldsymbol{\Sigma}}_T(\widehat{\gamma}_T)\right]_{11} = \frac{1}{T}\sum_{t=1}^{T}\hat{u}_t^2 \longrightarrow_p \mathbb{E}(u_t^2),
\tag{C.12}
$$

and

$$
\left[\widehat{\boldsymbol{\Gamma}}_T(\widehat{\gamma}_T, b_T)\right]_{11} = \sum_{i=1}^{T-1}k\left(\frac{i}{b_T}\right)\left(\frac{1}{T}\sum_{t=1}^{T-i}\hat{u}_{t+i}\hat{u}_t\right) \longrightarrow_p \lim_{T\to\infty}\frac{1}{T}\sum_{i=1}^{T-1}\sum_{t=1}^{T-i}\mathbb{E}(u_{t+i}u_t).
\tag{C.13}
$$

The consistency proofs for the other elements in the first row/column of these matrices follows easily using similar arguments. The following result will be used throughout

$$
\begin{aligned}
\hat{u}_t - u_t &= \boldsymbol{z}_t(\boldsymbol{\theta}_0)'\left[\begin{smallmatrix}\boldsymbol{\tau}_0\\ \boldsymbol{\phi}_0\end{smallmatrix}\right] - \boldsymbol{z}_t(\widehat{\boldsymbol{\theta}}_T)'\left[\begin{smallmatrix}\widehat{\boldsymbol{\tau}}_T\\ \widehat{\boldsymbol{\phi}}_T\end{smallmatrix}\right] \\
&= [\boldsymbol{z}_t(\boldsymbol{\theta}_0) - \boldsymbol{z}_t(\widehat{\boldsymbol{\theta}}_T)]'\left[\begin{smallmatrix}\boldsymbol{\tau}_0\\ \boldsymbol{\phi}_0\end{smallmatrix}\right] + [\boldsymbol{z}_t(\boldsymbol{\theta}_0) - \boldsymbol{z}_t(\widehat{\boldsymbol{\theta}}_T)]'\left[\begin{smallmatrix}\widehat{\boldsymbol{\tau}}_T-\boldsymbol{\tau}_0\\ \widehat{\boldsymbol{\phi}}_T-\boldsymbol{\phi}_0\end{smallmatrix}\right] - \boldsymbol{z}_t(\boldsymbol{\theta}_0)'\left[\begin{smallmatrix}\widehat{\boldsymbol{\tau}}_T-\boldsymbol{\tau}_0\\ \widehat{\boldsymbol{\phi}}_T-\boldsymbol{\phi}_0\end{smallmatrix}\right] \\
&= [\boldsymbol{d}_t(\boldsymbol{\theta}_0) - \boldsymbol{d}_t(\widehat{\boldsymbol{\theta}}_T)]'\boldsymbol{\tau}_0 + [\boldsymbol{d}_t(\boldsymbol{\theta}_0) - \boldsymbol{d}_t(\widehat{\boldsymbol{\theta}}_T)]'(\widehat{\boldsymbol{\tau}}_T - \boldsymbol{\tau}_0) - \boldsymbol{z}_t(\boldsymbol{\theta}_0)'\left[\begin{smallmatrix}\widehat{\boldsymbol{\tau}}_T-\boldsymbol{\tau}_0\\ \widehat{\boldsymbol{\phi}}_T-\boldsymbol{\phi}_0\end{smallmatrix}\right].
\end{aligned}
\tag{C.14}
$$

(i) We first show (C.12). Standard arguments provide $\frac{1}{T}\sum_{t=1}^{T} u_t^2 \longrightarrow_p \mathbb{E}(u_t^2)$, so it suffices to show that $\frac{1}{T}\sum_{t=1}^{T}(u_t^2 - \hat{u}_t^2) = o_p(1)$. First, by Cauchy-Schwarz we find

$$\begin{aligned}\left|\frac{1}{T}\sum_{t=1}^{T}(\hat{u}_t^2 - u_t^2)\right| &= \frac{1}{T}\sum_{t=1}^{T}(\hat{u}_t - u_t)^2 + \frac{2}{T}\left|\sum_{t=1}^{T} u_t(\hat{u}_t - u_t)\right| \\ &\leq \frac{1}{T}\sum_{t=1}^{T}(\hat{u}_t - u_t)^2 + 2\sqrt{\frac{1}{T}\sum_{t=1}^{T} u_t^2}\sqrt{\frac{1}{T}\sum_{t=1}^{T}(\hat{u}_t - u_t)^2}.\end{aligned} \tag{C.15}$$

It remains to establish $\frac{1}{T}\sum_{t=1}^{T}(\hat{u}_t - u_t)^2 = o_p(1)$. Using (C.14) and $\frac{1}{T}\sum_{t=1}^{T}(a_t + b_t)^2 \leq \frac{1}{T}\sum_{t=1}^{T} a_t^2 + \frac{1}{T}\sum_{t=1}^{T} b_t^2 + 2\sqrt{\frac{1}{T}\sum_{t=1}^{T} a_t^2}\sqrt{\frac{1}{T}\sum_{t=1}^{T} b_t^2}$, we see that the result follows if the following three statements are true:

$$\boldsymbol{\tau}_0'\left(\frac{1}{T}\sum_{t=1}^{T}[\boldsymbol{d}_t(\widehat{\boldsymbol{\theta}}_T) - \boldsymbol{d}_t(\boldsymbol{\theta}_0)][\boldsymbol{d}_t(\widehat{\boldsymbol{\theta}}_T) - \boldsymbol{d}_t(\boldsymbol{\theta}_0)]'\right)\boldsymbol{\tau}_0 = o_p(1), \tag{C.16a}$$

$$(\widehat{\boldsymbol{\tau}}_T - \boldsymbol{\tau}_0)'\left(\frac{1}{T}\sum_{t=1}^{T}[\boldsymbol{d}_t(\widehat{\boldsymbol{\theta}}_T) - \boldsymbol{d}_t(\boldsymbol{\theta}_0)][\boldsymbol{d}_t(\boldsymbol{\theta}_0) - \boldsymbol{d}_t(\widehat{\boldsymbol{\theta}}_T)]'\right)(\widehat{\boldsymbol{\tau}}_T - \boldsymbol{\tau}_0) = o_p(1), \tag{C.16b}$$

$$\begin{bmatrix} \widehat{\boldsymbol{\tau}}_T - \boldsymbol{\tau}_0 \\ \widehat{\boldsymbol{\phi}}_T - \boldsymbol{\phi}_0 \end{bmatrix}' \frac{1}{T}\sum_{t=1}^{T} \boldsymbol{z}_t(\boldsymbol{\theta}_0)\boldsymbol{z}_t(\boldsymbol{\theta}_0)' \begin{bmatrix} \widehat{\boldsymbol{\tau}}_T - \boldsymbol{\tau}_0 \\ \widehat{\boldsymbol{\phi}}_T - \boldsymbol{\phi}_0 \end{bmatrix} = o_p(1). \tag{C.16c}$$

We first look at the norm of the $(i, j)^{th}$ component of $\frac{1}{T}\sum_{t=1}^{T}[\boldsymbol{d}_t(\widehat{\boldsymbol{\theta}}_T) - \boldsymbol{d}_t(\boldsymbol{\theta}_0)][\boldsymbol{d}_t(\widehat{\boldsymbol{\theta}}_T) - \boldsymbol{d}_t(\boldsymbol{\theta}_0)]'$, i.e.

$$\begin{aligned}\left|\frac{1}{T}\sum_{t=1}^{T}\left(t^{\widehat{\theta}_i} - t^{\theta_{0i}}\right)\left(t^{\widehat{\theta}_j} - t^{\theta_{0j}}\right)\right| &= \frac{1}{T}\sum_{t=1}^{T} t^{\theta_{0i}+\theta_{0j}}\left|t^{\widehat{\theta}_i - \theta_{0i}} - 1\right|\left|t^{\widehat{\theta}_j - \theta_{0j}} - 1\right| \\ &\leq C\frac{1}{T}\sum_{t=1}^{T} t^{\theta_{0i}+\theta_{0j}}(\ln t)^2\left|\widehat{\theta}_i - \theta_{0i}\right|\left|\widehat{\theta}_j - \theta_{0j}\right| \\ &\leq C\frac{(\ln T)^2}{T}\left|T^{\theta_{0i}+\frac{1}{2}}(\widehat{\theta}_i - \theta_{0i})\right|\left|T^{\theta_{0j}+\frac{1}{2}}(\widehat{\theta}_j - \theta_{0j})\right|\frac{1}{T}\sum_{t=1}^{T}\left(\frac{t}{T}\right)^{\theta_{0i}+\theta_{0j}}.\end{aligned} \tag{C.17}$$

The RHS of (C.17) is $O_p\left(T^{-1}(\ln T)^2\right)$ by the convergence result from Theorem 1 and Lemma 1(i). The statements in (C.16a) and (C.16b) follow easily. We subsequently introduce the scaling matrix $\mathcal{D}_{\boldsymbol{\theta}_0,T} = \begin{bmatrix} \boldsymbol{D}_{d,T}(\boldsymbol{\theta}_0) & \mathbf{O} \\ \mathbf{O} & \boldsymbol{D}_{s,T} \end{bmatrix}$. The LHS of (C.16c) can now be expressed as

$$\left(\mathcal{D}_{\boldsymbol{\theta}_0,T}\begin{bmatrix} \widehat{\boldsymbol{\tau}}_T - \boldsymbol{\tau}_0 \\ \widehat{\boldsymbol{\phi}}_T - \boldsymbol{\phi}_0 \end{bmatrix}\right)'\left(\frac{1}{T}\sum_{t=1}^{T}[\mathcal{D}_{\boldsymbol{\theta}_0,T}^{-1}\boldsymbol{z}_t(\boldsymbol{\theta}_0)][\mathcal{D}_{\boldsymbol{\theta}_0,T}^{-1}\boldsymbol{z}_t(\boldsymbol{\theta}_0)]'\right)\left(\mathcal{D}_{\boldsymbol{\theta}_0,T}\begin{bmatrix} \widehat{\boldsymbol{\tau}}_T - \boldsymbol{\tau}_0 \\ \widehat{\boldsymbol{\phi}}_T - \boldsymbol{\phi}_0 \end{bmatrix}\right). \tag{C.18}$$

The results in Lemma 2(iii) imply that $\frac{1}{T}\sum_{t=1}^{T}[\mathcal{D}_{\boldsymbol{\theta}_0,T}^{-1}\boldsymbol{z}_t(\boldsymbol{\theta}_0)][\mathcal{D}_{\boldsymbol{\theta}_0,T}^{-1}\boldsymbol{z}_t(\boldsymbol{\theta}_0)]' \Rightarrow \int \widetilde{\boldsymbol{j}}(r;\boldsymbol{\theta}_0)\widetilde{\boldsymbol{j}}(r;\boldsymbol{\theta}_0)'dr$, where $\widetilde{\boldsymbol{j}}(r;\theta_0) = [\boldsymbol{d}(r;\boldsymbol{\theta}_0)', \boldsymbol{B}_{(1)}'(r), \ldots, \boldsymbol{B}_{(m)}'(r)]'$. A comparison of the elements of $\mathcal{D}_{\boldsymbol{\theta}_0,T}\begin{bmatrix} \widehat{\boldsymbol{\tau}}_T - \boldsymbol{\tau}_0 \\ \widehat{\boldsymbol{\phi}}_T - \boldsymbol{\phi}_0 \end{bmatrix}$ with the convergence rates of these estimators leads us to conclude that (C.16c) is also true.

C.3 Proof of (C.13)

(ii) To prove (C.13), we again show that the parameter estimation error is asymptotically negligible. If this holds, then the remainder of the proof follows from Jansson (2002). This said, we write

$$\begin{aligned}\sum_{i=1}^{T-1} k\left(\frac{i}{b_T}\right)\left(\frac{1}{T}\sum_{t=1}^{T-i}\left(\hat{u}_{t+i}\hat{u}_t - u_{t+i}u_t\right)\right) &= \sum_{i=1}^{T-1} k\left(\frac{i}{b_T}\right)\left(\frac{1}{T}\sum_{t=1}^{T-i} u_{t+i}(\hat{u}_t - u_t)\right)\\ &+ \sum_{i=1}^{T-1} k\left(\frac{i}{b_T}\right)\left(\frac{1}{T}\sum_{t=1}^{T-i}(\hat{u}_{t+i} - u_{t+i})u_t\right) + \sum_{i=1}^{T-1} k\left(\frac{i}{b_T}\right)\left(\frac{1}{T}\sum_{t=1}^{T-i}(\hat{u}_{t+i} - u_{t+i})(\hat{u}_t - u_t)\right) \\ &:= I + II + III.\end{aligned} \tag{C.19}$$

We provide details for $I = o_p(1)$ and omit the explicit proofs for II and III. Similar (and tedious) calculations are applicable there. Using (C.14), we can decompose I into

$$\begin{aligned}I = &\sum_{i=1}^{T-1} k\left(\frac{i}{b_T}\right)\left(\frac{1}{T}\sum_{t=1}^{T-i} u_{t+i}[\boldsymbol{d}_t(\boldsymbol{\theta}_0) - \boldsymbol{d}_t(\widehat{\boldsymbol{\theta}}_T)]'\boldsymbol{\tau}_0\right) + \sum_{i=1}^{T-1} k\left(\frac{i}{b_T}\right)\left(\frac{1}{T}\sum_{t=1}^{T-i} u_{t+i}[\boldsymbol{d}_t(\boldsymbol{\theta}_0) - \boldsymbol{d}_t(\widehat{\boldsymbol{\theta}}_T)]'\left(\widehat{\boldsymbol{\tau}}_T - \boldsymbol{\tau}_0\right)\right)\\ &- \sum_{i=1}^{T-1} k\left(\frac{i}{b_T}\right)\left(\frac{1}{T}\sum_{t=1}^{T-i} u_{t+i}\boldsymbol{z}_t(\boldsymbol{\theta}_0)'\left[\begin{smallmatrix}\widehat{\boldsymbol{\tau}}_T - \boldsymbol{\tau}_0\\ \widehat{\boldsymbol{\phi}}_T - \boldsymbol{\phi}_0\end{smallmatrix}\right]\right) := I_a + I_b - I_c.\end{aligned} \tag{C.20}$$

We adjust Hansen's (1992) argument slightly[19] and look at the quantities $\frac{T^{1/2}}{b_T \ln T}|I_i|$ for $i \in \{a, b, c\}$. If these quantities are stochastically bounded, then the result follows because $\frac{b_T \ln T}{T^{1/2}} \to 0$ by Assumption 3. By the Cauchy-Schwartz inequality, we have

$$\begin{aligned}\frac{T^{1/2}}{b_T \ln T}|I_a| &\leq \frac{T^{1/2}}{b_T \ln T}\sum_{i=1}^{T-1}\left|k\left(\frac{i}{b_T}\right)\right| \left|\frac{1}{T}\sum_{t=1}^{T-i} u_{t+i}[\boldsymbol{d}_t(\boldsymbol{\theta}_0) - \boldsymbol{d}_t(\widehat{\boldsymbol{\theta}}_T)]'\boldsymbol{\tau}_0\right|\\ &\leq \frac{T^{1/2}}{b_T \ln T}\sum_{i=1}^{T-i}\left|k\left(\frac{i}{b_T}\right)\right| \sqrt{\frac{1}{T}\sum_{t=1}^{T-i} u_{t+i}^2}\ \sqrt{\frac{1}{T}\sum_{t=1}^{T-i}\left([\boldsymbol{d}_t(\boldsymbol{\theta}_0) - \boldsymbol{d}_t(\widehat{\boldsymbol{\theta}}_T)]'\boldsymbol{\tau}_0\right)^2}\\ &\leq \frac{1}{b_T}\sum_{i=1}^{T-i}\left|k\left(\frac{i}{b_T}\right)\right| \sqrt{\frac{1}{T}\sum_{t=1}^{T} u_t^2}\ \sqrt{\boldsymbol{\tau}_0'\left(\frac{1}{(\ln T)^2}\sum_{t=1}^{T}[\boldsymbol{d}_t(\widehat{\boldsymbol{\theta}}_T) - \boldsymbol{d}_t(\boldsymbol{\theta}_0)][\boldsymbol{d}_t(\widehat{\boldsymbol{\theta}}_T) - \boldsymbol{d}_t(\boldsymbol{\theta}_0)]'\right)\boldsymbol{\tau}_0}\ .\end{aligned} \tag{C.21}$$

The RHS of (C.21) is bounded in probability due to lemma 1 of Jansson (2002), the fact that $\frac{1}{T}\sum_{t=1}^{T} u_t^2 \longrightarrow_p \mathbb{E}(u_t^2)$, and (C.17). Similarly, use

$$\begin{aligned}\frac{T^{1/2}}{b_T \ln T}|I_b| \leq &\frac{1}{b_T}\sum_{i=1}^{T-i}\left|k\left(\frac{i}{b_T}\right)\right| \sqrt{\frac{1}{T}\sum_{t=1}^{T} u_t^2}\ \times\\ &\sqrt{(\widehat{\boldsymbol{\tau}}_T - \boldsymbol{\tau}_0)'\left(\frac{1}{(\ln T)^2}\sum_{t=1}^{T}[\boldsymbol{d}_t(\widehat{\boldsymbol{\theta}}_T) - \boldsymbol{d}_t(\boldsymbol{\theta}_0)][\boldsymbol{d}_t(\widehat{\boldsymbol{\theta}}_T) - \boldsymbol{d}_t(\boldsymbol{\theta}_0)]'\right)(\widehat{\boldsymbol{\tau}}_T - \boldsymbol{\tau}_0)}\ .\end{aligned} \tag{C.22}$$

[19] Hansen (1992) multiplies his terms by $T^{1/2}/b_T$.

to show that $\frac{T^{1/2}}{b_T \ln T}|I_b| = O_p(1)$. Finally, we have

$$
\begin{aligned}
&\frac{T^{1/2}}{b_T \ln T}|I_c| \leq \frac{T^{1/2}}{b_T \ln T} \sum_{i=1}^{T-1} \left| k\left(\frac{i}{b_T}\right)\right| \left| \frac{1}{T}\sum_{t=1}^{T-i} u_{t+i} \boldsymbol{z}_t(\boldsymbol{\theta}_0)' \begin{bmatrix} \widehat{\boldsymbol{\tau}}_T - \boldsymbol{\tau}_0 \\ \widehat{\boldsymbol{\phi}}_T - \boldsymbol{\phi}_0 \end{bmatrix} \right| \\
&\leq \frac{1}{b_T} \sum_{i=1}^{T-i} \left| k\left(\frac{i}{b_T}\right)\right| \sqrt{\frac{1}{T}\sum_{t=1}^{T} u_t^2} \times \qquad \text{(C.23)} \\
&\sqrt{\left(\frac{\sqrt{T}}{\ln T}\mathcal{D}_{\boldsymbol{\theta}_0,T} \begin{bmatrix} \widehat{\boldsymbol{\tau}}_T - \boldsymbol{\tau}_0 \\ \widehat{\boldsymbol{\phi}}_T - \boldsymbol{\phi}_0 \end{bmatrix}\right)' \left(\frac{1}{T}\sum_{t=1}^{T} [\mathcal{D}_{\boldsymbol{\theta}_0,T}^{-1}\boldsymbol{z}_t(\boldsymbol{\theta}_0)][\mathcal{D}_{\boldsymbol{\theta}_0,T}^{-1}\boldsymbol{z}_t(\boldsymbol{\theta}_0)]'\right) \left(\frac{\sqrt{T}}{\ln T}\mathcal{D}_{\boldsymbol{\theta}_0,T} \begin{bmatrix} \widehat{\boldsymbol{\tau}}_T - \boldsymbol{\tau}_0 \\ \widehat{\boldsymbol{\phi}}_T - \boldsymbol{\phi}_0 \end{bmatrix}\right)}.
\end{aligned}
$$

Now note that $\frac{\sqrt{T}}{\ln T}\boldsymbol{D}_{d,T}(\boldsymbol{\theta}_0)(\widehat{\boldsymbol{\tau}}_T - \boldsymbol{\tau}_0)$ and $\sqrt{T}\boldsymbol{D}_{s,T}(\widehat{\boldsymbol{\phi}}_T - \boldsymbol{\phi}_0)$ are $O_p(1)$. This completes the proof. ■

C.4 Proof of Theorem 3

Proof of Theorem 3 For brevity, we define

$$
\widehat{\boldsymbol{\mathcal{J}}}_N\left(\widehat{\gamma}_T, \widehat{\boldsymbol{\Omega}}, \widehat{\boldsymbol{\Delta}}_{vu}^{-}\right) = \left\{ \boldsymbol{G}_{\widehat{\gamma},N}'^{-1} \left[\sum_{n=1}^{N} \dot{f}(\widehat{\boldsymbol{w}}_n, \widehat{\gamma}_T)\dot{f}(\widehat{\boldsymbol{w}}_n, \widehat{\gamma}_T)'\right] \boldsymbol{G}_{\widehat{\gamma},N}^{-1} \right\}^{-1} \left\{ \boldsymbol{G}_{\widehat{\gamma},N}'^{-1} \left[\sum_{n=1}^{N} \dot{f}(\widehat{\boldsymbol{w}}_n, \widehat{\gamma}_T)\widehat{\mu}_n\right] + \widehat{\boldsymbol{\mathcal{B}}}_{vu}^{-} \right\}.
$$

As a first step, we show $\widehat{\boldsymbol{\mathcal{J}}}_N\left(\widehat{\gamma}_T, \widehat{\boldsymbol{\Omega}}, \widehat{\boldsymbol{\Delta}}_{vu}^{-}\right) = \widehat{\boldsymbol{\mathcal{J}}}_N\left(\gamma_0, \widehat{\boldsymbol{\Omega}}, \widehat{\boldsymbol{\Delta}}_{vu}^{-}\right) + o_p^*(1)$. Direct calculation yields

$$
\begin{aligned}
\boldsymbol{R}_N &:= \boldsymbol{G}_{\gamma_0,N}\boldsymbol{G}_{\widehat{\gamma},N}^{-1} \\
&= \begin{bmatrix} \boldsymbol{D}_{d,N}(\boldsymbol{\theta}_0) & & \\ \boldsymbol{D}_{d,N}(\boldsymbol{\theta}_0)\,\mathrm{diag}[\boldsymbol{\tau}_0]\ln N & \boldsymbol{D}_{d,N}(\boldsymbol{\theta}_0) & \\ \mathbf{O}_{p\times d} & \mathbf{O}_{p\times d} & \boldsymbol{D}_{s,N} \end{bmatrix} \begin{bmatrix} \boldsymbol{D}_{d,N}(\widehat{\boldsymbol{\theta}}_T)^{-1} & & \\ -\boldsymbol{D}_{d,N}(\widehat{\boldsymbol{\theta}}_T)^{-1}\,\mathrm{diag}[\widehat{\boldsymbol{\tau}}_T]\ln N & \boldsymbol{D}_{d,N}(\widehat{\boldsymbol{\theta}}_T)^{-1} & \\ \mathbf{O}_{p\times d} & \mathbf{O}_{p\times d} & \boldsymbol{D}_{s,N}^{-1} \end{bmatrix} \\
&= \begin{bmatrix} \boldsymbol{D}_{d,N}(\boldsymbol{\theta}_0)\boldsymbol{D}_{d,N}(\widehat{\boldsymbol{\theta}}_T)^{-1} & & \\ \boldsymbol{D}_{d,N}(\boldsymbol{\theta}_0)\boldsymbol{D}_{d,N}(\widehat{\boldsymbol{\theta}}_T)^{-1}\,\mathrm{diag}[\boldsymbol{\tau}_0 - \widehat{\boldsymbol{\tau}}_T]\ln N & \boldsymbol{D}_{d,N}(\boldsymbol{\theta}_0)\boldsymbol{D}_{d,N}(\widehat{\boldsymbol{\theta}}_T)^{-1} & \\ \mathbf{O}_{p\times d} & \mathbf{O}_{p\times d} & \boldsymbol{I}_p \end{bmatrix}.
\end{aligned}
$$

We have $\boldsymbol{R}_N \longrightarrow_p \boldsymbol{I}_{2d+p}$. To see this, note that (1) a typical diagonal element of $\boldsymbol{D}_{d,N}(\boldsymbol{\theta}_0)\boldsymbol{D}_{d,N}(\widehat{\boldsymbol{\theta}}_T)^{-1}$ is $N^{\theta_{0i}-\widehat{\theta}_i}$ for which $N^{|\widehat{\theta}_i - \theta_{0i}|} = \exp\left(\left|\widehat{\theta}_i - \theta_{0i}\right| \ln N\right) \leq \exp\left((\ln N)N^{-(\theta_L+\frac{1}{2})}\left(\frac{N}{T}\right)^{\theta_L+\frac{1}{2}}\left|T^{\theta_{0i}+\frac{1}{2}}(\widehat{\theta}_i - \theta_{0i})\right|\right) \longrightarrow_p 1$, and (2) $\left\|\mathrm{diag}[\boldsymbol{\tau}_0 - \widehat{\boldsymbol{\tau}}_T]\ln N\right\| \leq \left\|\boldsymbol{D}_{d,T}(\boldsymbol{\theta}_0)^{-1}\right\| \frac{\ln T}{\sqrt{T}}(\ln N)\left\|\frac{\sqrt{T}}{\ln T}\boldsymbol{D}_{d,T}(\boldsymbol{\theta}_0)(\widehat{\boldsymbol{\tau}}_T - \boldsymbol{\tau}_0)\right\| \leq C\frac{\ln T}{T^{1/2+\theta_L}}(\ln N)O_p(1) = o_p(1)$. Define $\widetilde{\boldsymbol{\mathcal{N}}}_{\delta^*,N}(\gamma_0)$ similarly to $\widetilde{\boldsymbol{\mathcal{N}}}_{\delta,T}(\gamma_0)$ (page 39 below (C.1)). Consequently, if there exists a constant $\delta^* > 0$ such that $\mathbb{P}\left(\widehat{\gamma}_T \in \widetilde{\boldsymbol{\mathcal{N}}}_{\delta^*,N}(\gamma_0)\right) \to 1$, then

$$
\begin{aligned}
&\boldsymbol{G}_{\widehat{\gamma},N}'^{-1}\left[\sum_{n=1}^{N}\dot{f}(\widehat{\boldsymbol{w}}_n, \widehat{\gamma}_T)\dot{f}(\widehat{\boldsymbol{w}}_n, \widehat{\gamma}_T)'\right]\boldsymbol{G}_{\widehat{\gamma},N}^{-1} = \boldsymbol{R}_N'\boldsymbol{G}_{\gamma_0,N}'^{-1}\left[\sum_{n=1}^{N}\dot{f}(\widehat{\boldsymbol{w}}_n, \widehat{\gamma}_T)\dot{f}(\widehat{\boldsymbol{w}}_n, \widehat{\gamma}_T)'\right]\boldsymbol{G}_{\gamma_0,N}^{-1}\boldsymbol{R}_N \\
&= \boldsymbol{R}_N'\left\{\boldsymbol{G}_{\gamma_0,N}'^{-1}\left[\sum_{n=1}^{N}\dot{f}(\widehat{\boldsymbol{w}}_n, \gamma_0)\dot{f}(\widehat{\boldsymbol{w}}_n, \gamma_0)'\right]\boldsymbol{G}_{\gamma_0,N}^{-1} + o_p(1)\right\}\boldsymbol{R}_N \qquad \text{(C.24)} \\
&= \boldsymbol{G}_{\gamma_0,N}'^{-1}\left[\sum_{n=1}^{N}\dot{f}(\widehat{\boldsymbol{w}}_n, \gamma_0)\dot{f}(\widehat{\boldsymbol{w}}_n, \gamma_0)'\right]\boldsymbol{G}_{\gamma_0,N}^{-1} + o_p(1), \qquad \text{(C.25)}
\end{aligned}
$$

where (C.24) follows from the same arguments in (C.3). The condition $\mathbb{P}\left(\widehat{\gamma}_T \in \widetilde{\mathcal{N}}_{\delta^*,N}(\gamma_0)\right) \to 1$ is easily satisfied because $\left\|\boldsymbol{G}_{\gamma_0,N}(\widehat{\gamma}_T - \gamma_0)\right\| \leq \left\|\boldsymbol{G}_{\gamma_0,N}\boldsymbol{G}_{\gamma_0,T}^{-1}\right\| \left\|\boldsymbol{G}_{\gamma_0,T}(\widehat{\gamma}_T - \gamma_0)\right\| \leq C\delta \ln T =: \delta^* \ln T$, where

$$\boldsymbol{G}_{\gamma_0,N}\boldsymbol{G}_{\gamma_0,T}^{-1} = \sqrt{\frac{N}{T}}\begin{bmatrix} \boldsymbol{D}_{d,N}(\boldsymbol{\theta}_0)\boldsymbol{D}_{d,T}(\boldsymbol{\theta}_0)^{-1} & & \\ \boldsymbol{D}_{d,N}(\boldsymbol{\theta}_0)\boldsymbol{D}_{d,T}(\boldsymbol{\theta}_0)^{-1}\,\mathrm{diag}[\boldsymbol{\tau}_0]\left(\ln\frac{N}{T}\right) & \boldsymbol{D}_{d,N}(\boldsymbol{\theta}_0)\boldsymbol{D}_{d,T}(\boldsymbol{\theta}_0)^{-1} & \\ \mathbf{O}_{p\times d} & \mathbf{O}_{p\times d} & \boldsymbol{D}_{s,N}\boldsymbol{D}_{s,T}^{-1} \end{bmatrix},$$

and thus, by the norm property $\|\cdot\|^2 \leq \|\cdot\|_{\mathcal{F}}^2$,

$$\begin{aligned}
\left\|\boldsymbol{G}_{\gamma_0,N}\boldsymbol{G}_{\gamma_0,T}^{-1}\right\|^2 &\leq C\frac{N}{T}\left(\left\|\boldsymbol{D}_{d,N}(\boldsymbol{\theta}_0)\boldsymbol{D}_{d,T}^{-1}(\boldsymbol{\theta}_0)\right\|_{\mathcal{F}}^2\left(\ln\frac{N}{T}\right)^2 + \left\|\boldsymbol{D}_{s,N}\boldsymbol{D}_{s,T}^{-1}\right\|_{\mathcal{F}}^2\right) \\
&\leq C\left(\frac{N}{T}\right)^{1+2\theta_L}\left(\ln\frac{N}{T}\right)^2 + \left(\frac{N}{T}\right)^2 \leq C.
\end{aligned}$$

Next, we consider $\boldsymbol{G}_{\widehat{\gamma},N}^{\prime -1}\sum_{n=1}^N \dot{f}(\widehat{\boldsymbol{w}}_n, \widehat{\gamma}_T)\widehat{\mu}_n$, or equivalently

$$\boldsymbol{G}_{\widehat{\gamma},N}^{\prime -1}\left[\sum_{n=1}^N \dot{f}(\widehat{\boldsymbol{w}}_n, \widehat{\gamma}_T)\widehat{\mu}_n\right] = \boldsymbol{R}_N'\left[\boldsymbol{G}_{\gamma_0,N}^{\prime -1}\sum_{n=1}^N \dot{f}(\widehat{\boldsymbol{w}}_n, \gamma_0)\widehat{\mu}_n + \boldsymbol{G}_{\gamma_0,N}^{\prime -1}\sum_{n=1}^N\left(\dot{f}(\widehat{\boldsymbol{w}}_n, \widehat{\gamma}_T) - \dot{f}(\widehat{\boldsymbol{w}}_n, \gamma_0)\right)\widehat{\mu}_n\right].$$

We know $\boldsymbol{R}_N \longrightarrow_p \boldsymbol{I}_{2d+p}$. Moreover, by the triangle and Cauchy-Schwartz inequality, we have

$$\begin{aligned}
&\left\|\boldsymbol{G}_{\gamma_0,N}^{\prime -1}\sum_{n=1}^N\left(\dot{f}(\widehat{\boldsymbol{w}}_n, \widehat{\gamma}_T) - \dot{f}(\widehat{\boldsymbol{w}}_n, \gamma_0)\right)\widehat{\mu}_n\right\| \\
&\qquad \leq \sqrt{N\sum_{n=1}^N\left\|\boldsymbol{G}_{\gamma_0,N}^{\prime -1}\left(\dot{f}(\widehat{\boldsymbol{w}}_n, \widehat{\gamma}_T) - \dot{f}(\widehat{\boldsymbol{w}}_n, \gamma_0)\right)\right\|^2}\sqrt{\frac{1}{N}\sum_{n=1}^N \widehat{\mu}_n^2}.
\end{aligned} \tag{C.26}$$

Using $\boldsymbol{G}_{\gamma_0,N} = \boldsymbol{D}_{\boldsymbol{\theta}_0,N}\boldsymbol{L}_{\boldsymbol{\tau}_0,N}^{\prime -1}$, we see from (A.22) that the first term in the RHS of (C.26) does not depend on $\{\boldsymbol{e}_1, \ldots, \boldsymbol{e}_N\}$. As in (C.4), we also have $N\sum_{n=1}^N\left\|\boldsymbol{G}_{\gamma_0,N}^{\prime -1}\left(\dot{f}(\widehat{\boldsymbol{w}}_n, \widehat{\gamma}_T) - \dot{f}(\widehat{\boldsymbol{w}}_n, \gamma_0)\right)\right\|^2 \leq C\sum_{i=1}^4 I_i$ with

$$\begin{aligned}
I_1 &= \sum_{n=1}^N\left\|\boldsymbol{D}_{d,N}(\boldsymbol{\theta}_0)^{-1}\left((\widehat{\boldsymbol{\tau}}_T - \boldsymbol{\tau}_0)\odot(\boldsymbol{d}_n(\widehat{\boldsymbol{\theta}}_T) - \boldsymbol{d}_n(\boldsymbol{\theta}_0))\right)\ln n\right\|^2 = O_p\left(\frac{(\ln T)^6}{T^{1+2\theta_L}}\frac{N}{T^{1+2\theta_L}}\right) = o_p(1), \\
I_2 &= \sum_{n=1}^N\left\|\boldsymbol{D}_{d,N}(\boldsymbol{\theta}_0)^{-1}(\boldsymbol{d}_n(\widehat{\boldsymbol{\theta}}_T) - \boldsymbol{d}_n(\boldsymbol{\theta}_0))\right\|^2 = O_p\left((\ln T)^2\frac{N}{T^{1+2\theta_L}}\right) = o_p(1), \\
I_3 &= \sum_{n=1}^N\left\|\boldsymbol{D}_{d,N}(\boldsymbol{\theta}_0)^{-1}\left(\boldsymbol{\tau}_0\odot(\boldsymbol{d}_n(\widehat{\boldsymbol{\theta}}_T) - \boldsymbol{d}_n(\boldsymbol{\theta}_0))\right)\ln\frac{n}{N}\right\|^2 = O_p\left((\ln T)^4\frac{N}{T^{1+2\theta_L}}\right) = o_p(1), \\
I_4 &= \sum_{n=1}^N\left\|\boldsymbol{D}_{d,N}(\boldsymbol{\theta}_0)^{-1}\left((\widehat{\boldsymbol{\tau}}_T - \boldsymbol{\tau}_0)\odot\boldsymbol{d}_n(\boldsymbol{\theta}_0)\right)\ln n\right\|^2 = O_p\left((\ln T)^4\frac{N}{T^{1+2\theta_L}}\right) = o_p(1),
\end{aligned}$$

and where stochastic orders are established as in Lemma 4 (*i*) to (*iv*) (and thus omitted). For the second term in the RHS of (C.26), note that $\frac{1}{N}\sum_{n=1}^N \widehat{\mu}_n^2 \leq \frac{1}{N}\sum_{n=1}^N\left(\widehat{\mu}_n^2 + \widehat{\boldsymbol{v}}_n'\widehat{\boldsymbol{v}}_n\right) = \frac{1}{N}\sum_{n=1}^N \boldsymbol{e}_n'\widehat{\boldsymbol{\Omega}}_T\boldsymbol{e}_n \leq \left\|\widehat{\boldsymbol{\Omega}}_T\right\|\left(\frac{1}{N}\sum_{n=1}^N \boldsymbol{e}_n'\boldsymbol{e}_n\right) = O_p^*(1)$ because $\widehat{\boldsymbol{\Omega}}_T \longrightarrow_p \boldsymbol{\Omega}$ and $\boldsymbol{e}_n \overset{i.i.d.}{\sim} \mathrm{N}(\mathbf{0}, \boldsymbol{I}_{m+1})$. Overall, we have $\boldsymbol{G}_{\widehat{\gamma},N}^{\prime -1}\left[\sum_{n=1}^N \dot{f}(\widehat{\boldsymbol{w}}_n, \widehat{\gamma}_T)\widehat{\mu}_n\right] = \boldsymbol{G}_{\gamma_0,N}^{\prime -1}\sum_{n=1}^N \dot{f}(\widehat{\boldsymbol{w}}_n, \gamma_0)\widehat{\mu}_n + o_p^*(1)$. Combining this result with (C.25) gives $\widehat{\mathcal{J}}_N\left(\widehat{\gamma}_T, \widehat{\boldsymbol{\Omega}}, \widehat{\boldsymbol{\Delta}}_{vu}^-\right) = \widehat{\mathcal{J}}_N\left(\gamma_0, \widehat{\boldsymbol{\Omega}}, \widehat{\boldsymbol{\Delta}}_{vu}^-\right) + o_p^*(1)$.

Finally, we consider $\widehat{\mathcal{J}}_N\left(\gamma_0, \widehat{\boldsymbol{\Omega}}, \widehat{\boldsymbol{\Delta}}^-_{vu}\right)$ itself. By independence between $\{\boldsymbol{e}_n\}$ and $\{\widehat{\boldsymbol{\Omega}}, \widehat{\boldsymbol{\Delta}}^-_{vu}\}$, consistency of $\widehat{\boldsymbol{\Omega}}$, and a FCLT for the i.i.d. sequence, we have

$$\frac{1}{\sqrt{N}}\sum_{n=1}^{[rN]}\begin{bmatrix}\widehat{\mu}_n\\ \widehat{\boldsymbol{v}}_n\end{bmatrix} = \widehat{\boldsymbol{\Omega}}^{1/2}\frac{1}{\sqrt{N}}\sum_{n=1}^{[rN]}\boldsymbol{e}_n \xrightarrow{d^*} \boldsymbol{B}(r). \tag{C.27}$$

Note that the elements of $\widehat{\boldsymbol{\Omega}}$ and $\widehat{\boldsymbol{\Delta}}$ are always multiplicative in the construction. By virtue of (C.27) and the direct application of Lemma 2,

$$\begin{aligned}\boldsymbol{G}'^{-1}_{\gamma_0,N}\left[\sum_{n=1}^{N}\dot{f}(\widehat{\boldsymbol{w}}_n,\gamma_0)\widehat{\mu}_n\right] + \widehat{\mathcal{B}}^-_{vu} &\xrightarrow{d^*} \int \boldsymbol{j}(r;\gamma_0)dB_u(r) + \begin{bmatrix}\boldsymbol{0}_{d\times 1}\\ \boldsymbol{0}_{d\times 1}\\ \boldsymbol{\Omega}_{v_1u}\boldsymbol{b}_1\\ \vdots\\ \boldsymbol{\Omega}_{v_mu}\boldsymbol{b}_m\end{bmatrix} + \begin{bmatrix}\boldsymbol{0}_{d\times 1}\\ \boldsymbol{0}_{d\times 1}\\ \boldsymbol{\Delta}^-_{v_1u}\boldsymbol{b}_1\\ \vdots\\ \boldsymbol{\Delta}^-_{v_mu}\boldsymbol{b}_m\end{bmatrix}\\ &= \int_0^1 \boldsymbol{j}(r;\gamma_0)dB_u(r) + \begin{bmatrix}\boldsymbol{0}_{d\times 1}\\ \boldsymbol{0}_{d\times 1}\\ \boldsymbol{\Delta}_{v_1u}\boldsymbol{b}_1\\ \vdots\\ \boldsymbol{\Delta}_{v_mu}\boldsymbol{b}_m\end{bmatrix},\end{aligned} \tag{C.28}$$

where we use $\boldsymbol{\Omega} + \boldsymbol{\Delta}^- = (\boldsymbol{\Delta} + \boldsymbol{\Delta}' - \boldsymbol{\Sigma}) + (\boldsymbol{\Sigma} - \boldsymbol{\Delta}') = \boldsymbol{\Delta}$. Similarly, we have

$$\boldsymbol{G}'^{-1}_{\gamma_0,N}\left[\sum_{n=1}^{N}\dot{f}(\widehat{\boldsymbol{w}}_n,\gamma_0)\dot{f}(\widehat{\boldsymbol{w}}_n,\gamma_0)'\right]\boldsymbol{G}^{-1}_{\gamma_0,N} \xrightarrow{d^*} \int \boldsymbol{j}(r;\gamma_0)\boldsymbol{j}(r;\gamma_0)'dr. \tag{C.29}$$

By (C.28) and (C.29), we obtain the theorem. ■

D Proof of Theorem 4

Proof of Theorem 4 Without loss of generality, we set $\ell = 1$. Subsequently, note that

$$\begin{aligned}\frac{1}{\sqrt{q_T}}\sum_{t=1}^{[rq_T]}\hat{u}^+_t &= \frac{1}{\sqrt{q_T}}\sum_{t=1}^{[rq_T]}(u_t - \boldsymbol{\Omega}_{uv}\boldsymbol{\Omega}^{-1}_{vv}\boldsymbol{v}_t) + (\boldsymbol{\Omega}_{uv}\boldsymbol{\Omega}^{-1}_{vv} - \widehat{\boldsymbol{\Omega}}_{uv}\widehat{\boldsymbol{\Omega}}^{-1}_{vv})\frac{1}{\sqrt{q_T}}\sum_{t=1}^{[rq_T]}\boldsymbol{v}_t\\ &- \frac{1}{\sqrt{q_T}}\sum_{t=1}^{[rq_T]}\left(\boldsymbol{d}_t(\widehat{\boldsymbol{\theta}}_T)'\widehat{\boldsymbol{\tau}}_T - \boldsymbol{d}_t(\boldsymbol{\theta}_0)'\boldsymbol{\tau}_0\right) - \frac{1}{\sqrt{q_T}}\sum_{t=1}^{[rq_T]}\boldsymbol{s}'_t\left(\widehat{\boldsymbol{\phi}}_T - \boldsymbol{\phi}_0\right) =: VIa + VIb - VIc - VId.\end{aligned} \tag{D.1}$$

Assumption 2 justifies the use of a functional central limit theorem for linear processes, e.g. Phillips and Solo (1992). Therefore, $VIa \Rightarrow B_{u.v}(r)$ and $\Omega^{-1}_{u.v}\frac{1}{q_T}\sum_{t=1}^{q_T}\left(\frac{1}{\sqrt{q_T}}\sum_{i=\ell}^{t}(u_t - \boldsymbol{\Omega}_{uv}\boldsymbol{\Omega}^{-1}_{vv}\boldsymbol{v}_t)\right)^2 \Rightarrow \int[W(r)]^2dr$ by the continuous mapping theorem for functionals. Theorem 4 will thus follow if we can show that VIb, VIc and VId are asymptotically negligible.

Because Assumptions 1-3 are required to hold, Theorem 2 implies that $\widehat{\boldsymbol{\Omega}}_T \longrightarrow_p \boldsymbol{\Omega}$ (and hence $\|\boldsymbol{\Omega}_{uv}\boldsymbol{\Omega}_{vv}^{-1} - \widehat{\boldsymbol{\Omega}}_{uv}\widehat{\boldsymbol{\Omega}}_{vv}^{-1}\| \longrightarrow_p 0$). It follows that $VIb = o_p(1)$. We decompose VIc in three parts:

$$\begin{aligned} VIc &= \frac{1}{\sqrt{q_T}}\sum_{t=1}^{[rq_T]}\left(\boldsymbol{d}_t(\widehat{\boldsymbol{\theta}}_T) - \boldsymbol{d}_t(\boldsymbol{\theta}_0)\right)'\boldsymbol{\tau}_0 + \frac{1}{\sqrt{q_T}}\sum_{t=1}^{[rq_T]}\boldsymbol{d}_t(\boldsymbol{\theta}_0)'(\widehat{\boldsymbol{\tau}}_T - \boldsymbol{\tau}_0) \\ &\quad + \frac{1}{\sqrt{q_T}}\sum_{t=1}^{[rq_T]}\left(\boldsymbol{d}_t(\widehat{\boldsymbol{\theta}}_T) - \boldsymbol{d}_t(\boldsymbol{\theta}_0)\right)'(\widehat{\boldsymbol{\tau}}_T - \boldsymbol{\tau}_0) =: VIc^{(1)} + VIc^{(2)} + VIc^{(3)}. \end{aligned} \tag{D.2}$$

$|VIc^{(1)}| = \left|\frac{1}{\sqrt{q_T}}\sum_{t=1}^{[rq_T]}\left(\sum_{i=1}^{d}(t^{\widehat{\theta}_i} - t^{\theta_{0i}})\right)\tau_{0i}\right| \leq C(\ln q_T)\left(\frac{q_T}{T}\right)^{\theta_L+\frac{1}{2}}\sum_{i=1}^{d}|\tau_{0i}|\left|T^{\theta_{0i}+\frac{1}{2}}(\widehat{\theta}_i - \theta_{0i})\right|\frac{1}{q_T}\sum_{t=1}^{q_T}\left(\frac{t}{q_T}\right)^{\theta_{0i}} = (\ln q_T)\left(\frac{q_T}{T}\right)^{\theta_L+\frac{1}{2}}O_p(1)$ by the mean value theorem, Lemma 1(i), and $T^{\theta_{0i}+\frac{1}{2}}(\widehat{\theta}_i - \theta_{0i}) = O_p(1)$. By the Cauchy-Schwartz and triangle inequality, we have

$$\begin{aligned} \left|VIc^{(2)}\right| &= \left|\frac{1}{\sqrt{q_T}}\sum_{t=1}^{[rq_T]}\left[\boldsymbol{D}_{d,q_T}(\boldsymbol{\theta}_0)^{-1}\boldsymbol{d}_t(\boldsymbol{\theta}_0)\right]'\left[\boldsymbol{D}_{d,q_T}(\boldsymbol{\theta}_0)(\widehat{\boldsymbol{\tau}}_T - \boldsymbol{\tau}_0)\right]\right| \\ &= (\ln T)\left(\frac{q_T}{T}\right)^{\theta_L+\frac{1}{2}}\left(\frac{1}{q_T}\sum_{t=1}^{[rq_T]}\left\|\boldsymbol{D}_{d,q_T}(\boldsymbol{\theta}_0)^{-1}\boldsymbol{d}_t(\boldsymbol{\theta}_0)\right\|\right)\left\|\frac{\sqrt{T}}{\ln T}\boldsymbol{D}_{d,T}(\boldsymbol{\theta}_0)(\widehat{\boldsymbol{\tau}}_T - \boldsymbol{\tau}_0)\right\| = (\ln T)\left(\frac{q_T}{T}\right)^{\theta_L+\frac{1}{2}}O_p(1), \end{aligned}$$

where we used $\frac{\sqrt{T}}{\ln T}\boldsymbol{D}_{d,T}(\boldsymbol{\theta}_0)(\widehat{\boldsymbol{\tau}}_T - \boldsymbol{\tau}_0) = O_p(1)$ (see Theorem 1). Similarly, we bound

$$\left|VIc^{(3)}\right| \leq (\ln T)\left(\frac{q_T}{T}\right)^{\theta_L+\frac{1}{2}}\left(\frac{1}{q_T}\sum_{t=1}^{[rq_T]}\left\|\boldsymbol{D}_{d,q_T}(\boldsymbol{\theta}_0)^{-1}\left(\boldsymbol{d}_t(\widehat{\boldsymbol{\theta}}_T) - \boldsymbol{d}_t(\boldsymbol{\theta}_0)\right)\right\|\right)\left\|\frac{\sqrt{T}}{\ln T}\boldsymbol{D}_{d,T}(\boldsymbol{\theta}_0)(\widehat{\boldsymbol{\tau}}_T - \boldsymbol{\tau}_0)\right\|.$$

Because $\frac{1}{q_T}\sum_{t=1}^{[rq_T]}\left|q_T^{-\theta_{0i}}(t^{\widehat{\theta}_i} - t^{\theta_{0i}})\right| \leq C(\ln q_T)|\,\widehat{\theta}_i - \theta_{0i}|\left(\frac{1}{q_T}\sum_{t=1}^{[rq_T]}\left(\frac{t}{q_T}\right)^{\theta_{0i}}\right) = (\ln q_T)O_p\left(T^{-(\theta_L+\frac{1}{2})}\right)$ for any $i = 1, 2, \ldots, d$, we see that $VIc^{(3)} = o_p(1)$. Overall, $VIc^{(1)}$, $VIc^{(2)}$ and $VIc^{(3)}$ are all three asymptotically negligle under the prerequisite that $(\ln T)(q_T/T)^{\theta_L+\frac{1}{2}} \to 0$. Finally, term VId. From

$$\begin{aligned} |VId| &= \left|\left(\frac{q_T}{T}\right)^{1/2}\frac{1}{q_T}\sum_{t=1}^{[rq_T]}\left(\boldsymbol{D}_{s,q_T}^{-1}\boldsymbol{s}_t\right)'\boldsymbol{D}_{s,q_T}\boldsymbol{D}_{s,T}^{-1}\left[\sqrt{T}\boldsymbol{D}_{s,T}\left(\widehat{\boldsymbol{\phi}}_T - \boldsymbol{\phi}_0\right)\right]\right| \\ &\leq \left(\frac{q_T}{T}\right)\frac{1}{q_T}\sum_{t=1}^{[rq_T]}\left\|\boldsymbol{D}_{s,q_T}^{-1}\boldsymbol{s}_t\right\|\left\|\sqrt{T}\boldsymbol{D}_{s,T}\left(\widehat{\boldsymbol{\phi}}_T - \boldsymbol{\phi}_0\right)\right\| \leq \left(\frac{q_T}{T}\right)\left(\frac{1}{q_T}\sum_{t=1}^{[rq_T]}\left\|\boldsymbol{D}_{s,q_T}^{-1}\boldsymbol{s}_t\right\|^2\right)^{1/2}\left\|\sqrt{T}\boldsymbol{D}_{s,T}\left(\widehat{\boldsymbol{\phi}}_T - \boldsymbol{\phi}_0\right)\right\|, \end{aligned}$$

we see that $|VId| = O_p(\frac{q_T}{T})$ because: (1) $\frac{1}{q_T}\sum_{t=1}^{[rq_T]}\left\|\boldsymbol{D}_{s,q_T}^{-1}\boldsymbol{s}_t\right\|^2 = \mathrm{tr}\left(\frac{1}{q_T}\sum_{t=1}^{[rq_T]}\boldsymbol{D}_{s,q_T}^{-1}\boldsymbol{s}_t\boldsymbol{s}_t'\boldsymbol{D}_{s,q_T}^{-1}\right) = O_p(1)$, and (2) $\sqrt{T}\boldsymbol{D}_{s,T}\left(\widehat{\boldsymbol{\phi}}_T - \boldsymbol{\phi}_0\right) = O_p(1)$. ■

E Further Explanations for MC Results in the Introduction

The innovation sequences $\{u_t\}$ and $\{v_t\}$ are mutually independent and generated as $u_t \overset{i.i.d.}{\sim} \mathrm{N}(0, \sigma_u^2)$ and $v_t \overset{i.i.d.}{\sim} \mathrm{N}(0, \sigma_v^2)$. We used $\sigma_u^2 = \sigma_v^2 = 1$ for the simulations. The t-statistic is calculated as $t_{c_4=0} = \frac{\widehat{c}_4}{\sqrt{\widehat{\sigma}^2[\boldsymbol{X}'\boldsymbol{X}]_{44}^{-1}}}$, with $(T \times 4)$ matrix $\boldsymbol{X} = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ 1 & 2 & \cdots & T \\ x_1 & x_2 & \cdots & x_T \\ x_1^2 & x_2^2 & \cdots & x_T^2 \end{bmatrix}'$, $\widehat{c}_4$ is the third element of the OLS estimator $\widehat{\boldsymbol{c}} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y}$, and $\widehat{\sigma}^2 = \frac{1}{T}\left\|\boldsymbol{y} - \boldsymbol{X}\,\widehat{\boldsymbol{c}}\right\|^2$.[20] We reject the null whenever the test statistic is less than the 5% quantile of a standard normally distributed random variable.

We start with the derivation of the limiting distribution of $\widehat{\boldsymbol{c}}$. Define the scaling matrix $\boldsymbol{D}_T = \text{diag}(T^{1/2}, T^{3/2}, T, T^{3/2})$, such that

$$\boldsymbol{D}_T^{-1}\boldsymbol{X}'\boldsymbol{X}\boldsymbol{D}_T^{-1} = \begin{bmatrix} 1 & \frac{1}{T^2}\sum_{t=1}^T t & \frac{1}{T^{3/2}}\sum_{t=1}^T x_t & \frac{1}{T^2}\sum_{t=1}^T x_t^2 \\ * & \frac{1}{T^3}\sum_{t=1}^T t^2 & \frac{1}{T^{5/2}}\sum_{t=1}^T tx_t & \frac{1}{T^3}\sum_{t=1}^T tx_t^2 \\ * & * & \frac{1}{T^2}\sum_{t=1}^T x_t^2 & \frac{1}{T^{5/2}}\sum_{t=1}^T x_t^3 \\ * & * & * & \frac{1}{T^3}\sum_{t=1}^T x_t^4 \end{bmatrix} \Rightarrow \begin{bmatrix} 1 & \frac{1}{2} & \int B(r)dr & \int B^2(r)dr \\ * & \frac{1}{3} & \int rB(r)dr & \int rB^2(r)dr \\ * & * & \int B^2(r)dr & \int B^3(r)dr \\ * & * & * & \int B^4(r)dr \end{bmatrix}, \tag{E.1}$$

by Lemma 2(iii) and with $B(\cdot)$ denoting a Brownian motion such that $\mathbb{E}(B(s) - B(r))^2 = \sigma_v^2(s - r)$ for $(s > r)$. By the same lemma, after using the DGP $y_t = \tau_0 t^2 + u_t$, we have

$$\frac{1}{T^{5/2}}\boldsymbol{D}_T^{-1}\boldsymbol{X}'\boldsymbol{y} = \begin{bmatrix} \frac{1}{T^3}\sum_{t=1}^T y_t \\ \frac{1}{T^4}\sum_{t=1}^T ty_t \\ \frac{1}{T^{7/2}}\sum_{t=1}^T x_t y_t \\ \frac{1}{T^4}\sum_{t=1}^T x_t^2 y_t \end{bmatrix} = \tau_0 \begin{bmatrix} \frac{1}{T^3}\sum_{t=1}^T t^2 \\ \frac{1}{T^4}\sum_{t=1}^T t^3 \\ \frac{1}{T^{7/2}}\sum_{t=1}^T t^2 x_t \\ \frac{1}{T^4}\sum_{t=1}^T t^2 x_t^2 \end{bmatrix} + O_p(T^{-5/2}) \Rightarrow \tau_0 \begin{bmatrix} \frac{1}{3} \\ \frac{1}{4} \\ \int r^2 B(r)dr \\ \int r^2 B^2(r)dr \end{bmatrix}. \tag{E.2}$$

The combination (E.1) and (E.2) results in $\frac{1}{T^{5/2}}\boldsymbol{D}_T\widehat{\boldsymbol{c}} = O_p(1)$. For the fourth element this translates into $\frac{1}{T}\widehat{\boldsymbol{c}}_4 = O_p(1)$. Straightforward yet tedious calculations show $\frac{1}{T^4}\widehat{\sigma}^2 = O_p(1)$. The asymptotic behavior of the t-statistic is therefore

$$t_{c_4=0} = \frac{\widehat{c}_4}{\sqrt{\hat{\sigma}^2\frac{1}{T^3}[\boldsymbol{D}_T^{-1}\boldsymbol{X}'\boldsymbol{X}\boldsymbol{D}_T^{-1}]_{44}^{-1}}} = T^{1/2}\frac{\widehat{c}_4/T}{\sqrt{\frac{\hat{\sigma}^2}{T^4}[\boldsymbol{D}_T^{-1}\boldsymbol{X}'\boldsymbol{X}\boldsymbol{D}_T^{-1}]_{44}^{-1}}} = O_p(T^{1/2}). \tag{E.3}$$

The equation above shows that the t-statistic is stochastically unbounded whenever $\tau_0 \neq 0$.[21] Its sign is governed by τ_0, see (E.2).

F Limiting Distribution for Example 1

Referring to Theorem 1, we find

$$\begin{bmatrix} T^{\theta_0+\frac{1}{2}} & \\ T^{\theta_0+\frac{1}{2}}\tau_0\ln(T) & T^{\theta_0+\frac{1}{2}} \end{bmatrix}\begin{bmatrix} \widehat{\theta}_T - \theta_0 \\ \widehat{\tau}_T - \tau_0 \end{bmatrix} \Rightarrow \begin{bmatrix} \int(\tau_0 r^{\theta_0}\ln(r))^2 dr & \int \tau_0 r^{2\theta_0}\ln(r)dr \\ \int \tau_0 r^{2\theta_0}\ln(r)dr & \int r^{2\theta_0}dr \end{bmatrix}^{-1}\begin{bmatrix} \int \tau_0 r^{\theta_0}\ln(r)dB_u \\ \int r^{\theta_0}dB_u \end{bmatrix}.$$

We have to show that the quantity in the RHS is normally distributed with mean and variance as provided in (3.2) of the main paper. Consider an arbitrary vector $\boldsymbol{c} = [c_1, c_2]'$ and define

$$A_{\boldsymbol{c}} = \boldsymbol{c}'\begin{bmatrix} \int \tau_0 r^{\theta_0}\ln(r)dB_u \\ \int r^{\theta_0}dB_u \end{bmatrix} = \int\left[c_1\tau_0 r^{\theta_0}\ln(r) + c_2 r^{\theta_0}\right]dB_u \overset{d}{=} \Omega_{uu}^{1/2}\int\left[c_1\tau_0 r^{\theta_0}\ln(r) + c_2 r^{\theta_0}\right]dW_u.$$

[20]The t-statistic should be adjusted in the presence of serial correlation and/or endogeneity. For simplicity, we did not include such effects in the MC simulations. Similarly, our estimator for σ_u^2 also exploits the fact that $\{u_t\}$ is an i.i.d. sequence.

[21]If τ_0 equals zero, then $y_t = u_t$ and the terms that currently dominate the asymptotic distribution will be absent. The t-statistic will be asymptotically standard normally distributed.

Gaussianity is preserved under mean square integration (e.g. section 4.6 in Soong (1973)), so $A_{\boldsymbol{c}}$ it suffices to derive mean and variance. Equation (4.190) in the same reference yields $\mathbb{E}(A_{\boldsymbol{c}}) = \Omega_{uu}^{1/2}\int\left[c_1\tau_0 r^{\theta_0}\ln(r) + c_2 r^{\theta_0}\right]d\mathbb{E}(W_u) = 0$. Moreover, by (2.16) in Tanaka (2017)

$$\mathbb{V}\mathrm{ar}\left(A_{\boldsymbol{c}}\right) = \Omega_{uu}\int\left[c_1\tau_0 r^{\theta_0}\ln(r) + c_2 r^{\theta_0}\right]^2 dr = \Omega_{uu}\boldsymbol{c}'\begin{bmatrix}\int\left(\tau_0 r^{\theta_0}\ln(r)\right)^2 dr & \int\tau_0 r^{2\theta_0}\ln(r)dr\\ \int\tau_0 r^{2\theta_0}\ln(r)dr & \int r^{2\theta_0}dr\end{bmatrix}\boldsymbol{c}.$$

$\boldsymbol{c}$ was arbitrary and therefore $\begin{bmatrix}\int\tau_0 r^{\theta_0}\ln(r)dB_u\\ \int r^{\theta_0}dB_u\end{bmatrix}\sim \mathrm{N}\left(\mathbf{0},\Omega_{uu}\begin{bmatrix}\int\left(\tau_0 r^{\theta_0}\ln(r)\right)^2 dr & \int\tau_0 r^{2\theta_0}\ln(r)dr\\ \int\tau_0 r^{2\theta_0}\ln(r)dr & \int r^{2\theta_0}dr\end{bmatrix}\right)$. Finally, use $\int\left(r^{\theta_0}\ln(r)\right)^2 dr = \frac{2}{(2\theta_0+1)^3}$, $\int r^{2\theta_0}\ln(r)dr = -\frac{1}{(2\theta_0+1)^2}$, and basic linear algebra to recover the claim of Example 1.

G FMOLS Estimator

We here comment on the asymptotic properties of the FMOLS estimator. To be specific, we analyse the asymptotic properties of $\widetilde{\boldsymbol{D}}_{\boldsymbol{\theta}_0,T}\begin{bmatrix}\widehat{\boldsymbol{\tau}}_T^+-\boldsymbol{\tau}_0\\ \widehat{\boldsymbol{\phi}}_T^+-\boldsymbol{\phi}_0\end{bmatrix}$, with $\widetilde{\boldsymbol{D}}_{\boldsymbol{\theta}_0,T} = \sqrt{T}\begin{bmatrix}\boldsymbol{D}_{d,T}(\boldsymbol{\theta}_0) & \mathbf{O}_{d\times p}\\ \mathbf{O}_{p\times d} & \boldsymbol{D}_{s,T}\end{bmatrix}$ and

$$\begin{bmatrix}\widehat{\boldsymbol{\tau}}_T^+\\ \widehat{\boldsymbol{\phi}}_T^+\end{bmatrix} = \left(\sum_{t=1}^T \boldsymbol{z}_t(\widehat{\boldsymbol{\theta}}_T)\boldsymbol{z}_t(\widehat{\boldsymbol{\theta}}_T)'\right)^{-1}\left(\sum_{t=1}^T \boldsymbol{z}_t(\widehat{\boldsymbol{\theta}}_T)y_t^+ - \boldsymbol{A}^*\right),$$

where y_t^+ and $\boldsymbol{A}^*$ are the usual second-order bias corrections. That is, $y_t^+ = y_t - \widehat{\boldsymbol{\Omega}}_{uv}\widehat{\boldsymbol{\Omega}}_{vv}^{-1}\Delta\boldsymbol{x}_t$ and $\boldsymbol{A}^* = [\mathbf{0}_{d\times 1}', \boldsymbol{A}_1^{*\prime},\dots,\boldsymbol{A}_m^{*\prime}]'$ with $\boldsymbol{A}_i^* = \widehat{\boldsymbol{\Delta}}_{v_i u}^+\left[T, 2\sum_{t=1}^T x_{it},\dots,p_i\sum_{t=1}^T x_{it}^{p_i-1}\right]'$ and $\widehat{\boldsymbol{\Delta}}_{v_i u}^+$ is the i^{th} row of $\widehat{\boldsymbol{\Delta}}_{vu}^+ = \widehat{\boldsymbol{\Delta}}_{vu} - \widehat{\boldsymbol{\Delta}}_{vv}\widehat{\boldsymbol{\Omega}}_{vv}^{-1}\widehat{\boldsymbol{\Omega}}_{vu}$ ($i = 1,2,\dots,m$). If the converge speed of $\widehat{\boldsymbol{\theta}}_T$ is sufficiently fast, then its estimation error is asymptotically negligible and the limiting distribution of $\widetilde{\boldsymbol{D}}_{\boldsymbol{\theta}_0,T}\begin{bmatrix}\widehat{\boldsymbol{\tau}}_T^+-\boldsymbol{\tau}_0\\ \widehat{\boldsymbol{\phi}}_T^+-\boldsymbol{\phi}_0\end{bmatrix}$ is mixed normal.

We now focus on the limiting distribution. By linear algebra manipulations, we find

$$\widetilde{\boldsymbol{D}}_{\boldsymbol{\theta}_0,T}\begin{bmatrix}\widehat{\boldsymbol{\tau}}_T^+ - \boldsymbol{\tau}_0\\ \widehat{\boldsymbol{\phi}}_T^+ - \boldsymbol{\phi}_0\end{bmatrix} = \left(\widetilde{\boldsymbol{D}}_{\boldsymbol{\theta}_0,T}^{-1}\sum_{t=1}^T \boldsymbol{z}_t(\widehat{\boldsymbol{\theta}}_T)\boldsymbol{z}_t(\widehat{\boldsymbol{\theta}}_T)'\widetilde{\boldsymbol{D}}_{\boldsymbol{\theta}_0,T}^{-1}\right)^{-1}\widetilde{\boldsymbol{D}}_{\boldsymbol{\theta}_0,T}^{-1}\left[\sum_{t=1}^T \boldsymbol{z}_t(\widehat{\boldsymbol{\theta}}_T)\tilde{u}_t^+ - \boldsymbol{A}^*\right], \tag{G.1}$$

where $\tilde{u}_t^+ = \left(\boldsymbol{z}_t(\boldsymbol{\theta}_0)-\boldsymbol{z}_t(\widehat{\boldsymbol{\theta}}_T)\right)'\begin{bmatrix}\boldsymbol{\tau}_0\\ \boldsymbol{\phi}_0\end{bmatrix}+u_t-\widehat{\boldsymbol{\Omega}}_{uv}\widehat{\boldsymbol{\Omega}}_{vv}^{-1}\Delta\boldsymbol{x}_t$. We will discuss $\widetilde{\boldsymbol{D}}_{\boldsymbol{\theta}_0,T}^{-1}\sum_{t=1}^T \boldsymbol{z}_t(\widehat{\boldsymbol{\theta}}_T)\boldsymbol{z}_t(\widehat{\boldsymbol{\theta}}_T)'\widetilde{\boldsymbol{D}}_{\boldsymbol{\theta}_0,T}^{-1}$ and $\widetilde{\boldsymbol{D}}_{\boldsymbol{\theta}_0,T}^{-1}\left[\sum_{t=1}^T \boldsymbol{z}_t(\widehat{\boldsymbol{\theta}}_T)\tilde{u}_t^+ - \boldsymbol{A}^*\right]$ separately after having enumerate several intermediate results.

Lemma 5
Define $\widetilde{\boldsymbol{j}}(r;\boldsymbol{\theta}_0) = [\boldsymbol{d}(r;\boldsymbol{\theta}_0)', \boldsymbol{B}_{(1)}'(r),\dots,\boldsymbol{B}_{(m)}'(r)]'$ and $B_{u.v} = B_u - \boldsymbol{\Omega}_{uv}\boldsymbol{\Omega}_{vv}^{-1}\boldsymbol{B}_v$. Then, under Assumptions 1-3, we have

(i) $\widetilde{\boldsymbol{D}}_{\boldsymbol{\theta}_0,T}^{-1}\sum_{t=1}^T \boldsymbol{z}_t(\widehat{\boldsymbol{\theta}}_T)\boldsymbol{z}_t(\widehat{\boldsymbol{\theta}}_T)'\widetilde{\boldsymbol{D}}_{\boldsymbol{\theta}_0,T}^{-1} \Rightarrow \int\widetilde{\boldsymbol{j}}(r;\boldsymbol{\theta}_0)\widetilde{\boldsymbol{j}}(r;\boldsymbol{\theta}_0)'dr$,

(ii) $\widetilde{\boldsymbol{D}}_{\boldsymbol{\theta}_0,T}^{-1}\left[\sum_{t=1}^T \boldsymbol{z}_t(\boldsymbol{\theta}_0)\left(u_t - \widehat{\boldsymbol{\Omega}}_{uv}\widehat{\boldsymbol{\Omega}}_{vv}^{-1}\boldsymbol{v}_t\right) - \boldsymbol{A}^*\right] \Rightarrow \int\widetilde{\boldsymbol{j}}(r;\boldsymbol{\theta}_0)dB_{u.v}(r)$,

(iii) $\widetilde{\boldsymbol{D}}_{\boldsymbol{\theta}_0,T}^{-1}\sum_{t=1}^T \boldsymbol{z}_t(\boldsymbol{\theta}_0)\left(\boldsymbol{z}_t(\widehat{\boldsymbol{\theta}}_T) - \boldsymbol{z}_t(\boldsymbol{\theta}_0)\right)'\begin{bmatrix}\boldsymbol{\tau}_0\\ \boldsymbol{\phi}_0\end{bmatrix} = O_p\left(\ln T\right)$,

(iv) $\sum_{t=1}^{b_T}\widetilde{\boldsymbol{D}}_{\boldsymbol{\theta}_0,b_T}^{-1}\left(\boldsymbol{z}_t(\widehat{\boldsymbol{\theta}}_T) - \boldsymbol{z}_t(\boldsymbol{\theta}_0)\right)\left(\boldsymbol{z}_t(\widehat{\boldsymbol{\theta}}_T) - \boldsymbol{z}_t(\boldsymbol{\theta}_0)\right)'\begin{bmatrix}\boldsymbol{\tau}_0\\ \boldsymbol{\phi}_0\end{bmatrix} = O_p\left((\ln T)^2 T^{-(\theta_L+\frac{1}{2})}\right)$,

(v) $\widetilde{\boldsymbol{D}}_{\boldsymbol{\theta}_0,T}^{-1}\sum_{t=1}^T\left(\boldsymbol{z}_t(\widehat{\boldsymbol{\theta}}_T) - \boldsymbol{z}_t(\boldsymbol{\theta}_0)\right)\left(u_t - \widehat{\boldsymbol{\Omega}}_{uv}\widehat{\boldsymbol{\Omega}}_{vv}^{-1}\boldsymbol{v}_t\right) = o_p(1)$.

Proof ***(i)*** We can always add and subtract such that the LHS of (*i*) reads

$$\begin{aligned}\widetilde{\boldsymbol{D}}_{\boldsymbol{\theta}_0,T}^{-1}\sum_{t=1}^{T}\boldsymbol{z}_t(\widehat{\boldsymbol{\theta}}_T)\boldsymbol{z}_t(\widehat{\boldsymbol{\theta}}_T)'\widetilde{\boldsymbol{D}}_{\boldsymbol{\theta}_0,T} &= \widetilde{\boldsymbol{D}}_{\boldsymbol{\theta}_0,T}^{-1}\sum_{t=1}^{T}\boldsymbol{z}_t(\boldsymbol{\theta}_0)\boldsymbol{z}_t(\boldsymbol{\theta}_0)'\widetilde{\boldsymbol{D}}_{\boldsymbol{\theta}_0,T}^{-1}\\ &\quad+\left(\widetilde{\boldsymbol{D}}_{\boldsymbol{\theta}_0,T}^{-1}\sum_{t=1}^{T}\boldsymbol{z}_t(\widehat{\boldsymbol{\theta}}_T)\boldsymbol{z}_t(\widehat{\boldsymbol{\theta}}_T)'\widetilde{\boldsymbol{D}}_{\boldsymbol{\theta}_0,T}^{-1}-\widetilde{\boldsymbol{D}}_{\boldsymbol{\theta}_0,T}^{-1}\sum_{t=1}^{T}\boldsymbol{z}_t(\boldsymbol{\theta}_0)\boldsymbol{z}_t(\boldsymbol{\theta}_0)'\widetilde{\boldsymbol{D}}_{\boldsymbol{\theta}_0,T}^{-1}\right).\end{aligned} \tag{G.2}$$

Lemma 2(iii) implies that the first term in the RHS of (G.2) converges to $\int\widetilde{\boldsymbol{j}}(r;\boldsymbol{\theta}_0)\widetilde{\boldsymbol{j}}(r;\boldsymbol{\theta}_0)'dr$. It remains to show that the term in brackets vanishes. By $\sum_t\boldsymbol{a}_t\boldsymbol{a}_t'-\sum_t\boldsymbol{b}_t\boldsymbol{b}_t'=\sum_t(\boldsymbol{a}_t-\boldsymbol{b}_t)(\boldsymbol{a}_t-\boldsymbol{b}_t)'+\sum_t(\boldsymbol{a}_t-\boldsymbol{b}_t)\boldsymbol{b}_t'+\sum_t\boldsymbol{b}_t(\boldsymbol{a}_t-\boldsymbol{b}_t)'$ and the Cauchy-Schwarz inequality, we have

$$\begin{aligned}&\left\|\widetilde{\boldsymbol{D}}_{\boldsymbol{\theta}_0,T}^{-1}\sum_{t=1}^{T}\boldsymbol{z}_t(\widehat{\boldsymbol{\theta}}_T)\boldsymbol{z}_t(\widehat{\boldsymbol{\theta}}_T)'\widetilde{\boldsymbol{D}}_{\boldsymbol{\theta}_0,T}^{-1}-\widetilde{\boldsymbol{D}}_{\boldsymbol{\theta}_0,T}^{-1}\sum_{t=1}^{T}\boldsymbol{z}_t(\boldsymbol{\theta}_0)\boldsymbol{z}_t(\boldsymbol{\theta}_0)'\widetilde{\boldsymbol{D}}_{\boldsymbol{\theta}_0,T}^{-1}\right\|\\ &\le\sum_{t=1}^{T}\left\|\widetilde{\boldsymbol{D}}_{\boldsymbol{\theta}_0,T}^{-1}\left(\boldsymbol{z}_t(\widehat{\boldsymbol{\theta}}_T)-\boldsymbol{z}_t(\boldsymbol{\theta}_0)\right)\right\|^2+2\sum_{t=1}^{T}\left\|\widetilde{\boldsymbol{D}}_{\boldsymbol{\theta}_0,T}^{-1}\boldsymbol{z}_t(\boldsymbol{\theta}_0)\right\|\left\|\widetilde{\boldsymbol{D}}_{\boldsymbol{\theta}_0,T}^{-1}\left(\boldsymbol{z}_t(\widehat{\boldsymbol{\theta}}_T)-\boldsymbol{z}_t(\boldsymbol{\theta}_0)\right)\right\|\\ &\le\sum_{t=1}^{T}\left\|\widetilde{\boldsymbol{D}}_{\boldsymbol{\theta}_0,T}^{-1}\left(\boldsymbol{z}_t(\widehat{\boldsymbol{\theta}}_T)-\boldsymbol{z}_t(\boldsymbol{\theta}_0)\right)\right\|^2+2\sqrt{\sum_{t=1}^{T}\left\|\widetilde{\boldsymbol{D}}_{\boldsymbol{\theta}_0,T}^{-1}\boldsymbol{z}_t(\boldsymbol{\theta}_0)\right\|^2}\sqrt{\sum_{t=1}^{T}\left\|\widetilde{\boldsymbol{D}}_{\boldsymbol{\theta}_0,T}^{-1}\left(\boldsymbol{z}_t(\widehat{\boldsymbol{\theta}}_T)-\boldsymbol{z}_t(\boldsymbol{\theta}_0)\right)\right\|^2}.\end{aligned}$$

We have $\sum_{t=1}^{T}\left\|\widetilde{\boldsymbol{D}}_{\boldsymbol{\theta}_0,T}^{-1}\boldsymbol{z}_t(\boldsymbol{\theta}_0)\right\|^2=\operatorname{tr}\left(\sum_{t=1}^{T}\widetilde{\boldsymbol{D}}_{\boldsymbol{\theta}_0,T}^{-1}\boldsymbol{z}_t(\boldsymbol{\theta}_0)\boldsymbol{z}_t(\boldsymbol{\theta}_0)'\widetilde{\boldsymbol{D}}_{\boldsymbol{\theta}_0,T}^{-1}\right)\Rightarrow\operatorname{tr}\left(\int\widetilde{\boldsymbol{j}}(r;\boldsymbol{\theta}_0)\widetilde{\boldsymbol{j}}(r;\boldsymbol{\theta}_0)'dr\right)$. Next note that $\sum_{t=1}^{T}\left\|\widetilde{\boldsymbol{D}}_{\boldsymbol{\theta}_0,T}^{-1}(\boldsymbol{z}_t(\widehat{\boldsymbol{\theta}}_T)-\boldsymbol{z}_t(\boldsymbol{\theta}_0))\right\|^2=\frac{1}{T}\sum_{t=1}^{T}\left\|\boldsymbol{D}_{d,T}(\boldsymbol{\theta}_0)^{-1}(\boldsymbol{d}_t(\widehat{\boldsymbol{\theta}}_T)-\boldsymbol{d}_t(\boldsymbol{\theta}_0))\right\|^2$. A typical contribution to the latter sum of norms can be bound by

$$\begin{aligned}\frac{1}{T}\sum_{t=1}^{T}[T^{-\theta_{0i}}(t^{\widehat{\theta}_i}-t^{\theta_{0i}})]^2&\le C\left(\widehat{\theta}_i-\theta_{0i}\right)\frac{1}{T}\sum_{t=1}^{T}\left(\frac{t}{T}\right)^{2\theta_{0i}}(\ln t)^2\\ &\le CT^{-2(\theta_{0i}-\frac{1}{2})}(\ln T)^2\left[T^{\theta_{0i}+\frac{1}{2}}\left(\widehat{\theta}_i-\theta_{0i}\right)\right]^2\sup_{\theta_L\le\theta\le\theta_U}\left|\frac{1}{T}\sum_{t=1}^{T}\left(\frac{t}{T}\right)^{2\theta}\right|\le T^{-2(\theta_L+\frac{1}{2})}(\ln b_T)^2O_p(1)=o_p(1),\end{aligned} \tag{G.3}$$

where we used the mean-value theorem and Lemma 1(i). The claim follows. ***(ii)*** $\widehat{\boldsymbol{\Omega}}_{uv}$ and $\widehat{\boldsymbol{\Omega}}_{vv}$ are consistently estimating $\boldsymbol{\Omega}_{uv}$ and $\boldsymbol{\Omega}_{vv}$, respectively (Theorem 2). It therefore suffices to look at $\widetilde{\boldsymbol{D}}_{\boldsymbol{\theta}_0,T}^{-1}\sum_{t=1}^{T}\boldsymbol{z}_t(\boldsymbol{\theta}_0)\left(u_t-\boldsymbol{\Omega}_{uv}\boldsymbol{\Omega}_{vv}^{-1}\boldsymbol{v}_t\right)$ and $\widetilde{\boldsymbol{D}}_{\boldsymbol{\theta}_0,T}^{-1}\boldsymbol{A}^*$. Lemma 2(ii) with $u_t^+=u_t-\boldsymbol{\Omega}_{uv}\boldsymbol{\Omega}_{vv}^{-1}\boldsymbol{v}_t$ instead of u_t gives the limiting result $\frac{1}{\sqrt{T}}\sum_{t=1}^{T}(x_{it}/\sqrt{T})^ju_t^+\Rightarrow\int_0^1\boldsymbol{B}_{v_i}^j(r)dB_{u.v}(r)+j\boldsymbol{\Delta}_{v_iu}^+\int_0^1\boldsymbol{B}_{v_i}^{j-1}(r)dr$, which implies

$$\widetilde{\boldsymbol{D}}_{\boldsymbol{\theta}_0,T}^{-1}\sum_{t=1}^{T}\boldsymbol{z}_t(\boldsymbol{\theta}_0)\left(u_t-\boldsymbol{\Omega}_{uv}\boldsymbol{\Omega}_{vv}^{-1}\boldsymbol{v}_t\right)\Rightarrow\int\widetilde{\boldsymbol{j}}(r;\gamma_0)dB_{u.v}(r)+\widetilde{\mathscr{B}}_{vu}^+, \tag{G.4}$$

where $\widetilde{\mathscr{B}}_{vu}^+=[0_{d\times1}',\boldsymbol{b}_1'\boldsymbol{\Delta}_{v_1u}^+,\ldots,\boldsymbol{b}_m'\boldsymbol{\Delta}_{v_mu}^+]'$. The term $-\widetilde{\boldsymbol{D}}_{\boldsymbol{\theta}_0,T}^{-1}\boldsymbol{A}^*$ is constructed to asymptotically cancel out the term $\widetilde{\mathscr{B}}_{vu}^+$ in the RHS of (G.4). ***(iii)*** Using $\boldsymbol{z}_t(\widehat{\boldsymbol{\theta}}_T)-\boldsymbol{z}_t(\boldsymbol{\theta}_0)=\left[\begin{smallmatrix}\boldsymbol{d}_t(\widehat{\boldsymbol{\theta}}_T)-\boldsymbol{d}_t(\boldsymbol{\theta}_0)\\ \boldsymbol{0}\end{smallmatrix}\right]$, we have

$$\widetilde{\boldsymbol{D}}_{\boldsymbol{\theta}_0,T}^{-1}\sum_{t=1}^{T}\boldsymbol{z}_t(\boldsymbol{\theta}_0)\left(\boldsymbol{z}_t(\widehat{\boldsymbol{\theta}}_T)-\boldsymbol{z}_t(\boldsymbol{\theta}_0)\right)'\left[\begin{smallmatrix}\boldsymbol{\tau}_0\\ \boldsymbol{\phi}_0\end{smallmatrix}\right]=\begin{bmatrix}\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\boldsymbol{D}_{d,T}(\boldsymbol{\theta}_0)^{-1}\boldsymbol{d}_t(\boldsymbol{\theta}_0)\left(\boldsymbol{d}_t(\widehat{\boldsymbol{\theta}}_T)-\boldsymbol{d}_t(\boldsymbol{\theta}_0)\right)'\boldsymbol{\tau}_0\\ \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\boldsymbol{D}_{s,T}^{-1}\boldsymbol{s}_t\left(\boldsymbol{d}_t(\widehat{\boldsymbol{\theta}}_T)-\boldsymbol{d}_t(\boldsymbol{\theta}_0)\right)'\boldsymbol{\tau}_0\end{bmatrix}.$$

The typical elements in the vector on the RHS are of the form $\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left(\frac{t}{T}\right)^{\theta_{0i}}\sum_{k=1}^{d}\tau_{0k}(t^{\widehat{\theta}_k}-t^{\theta_{0k}})$ or $\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left(\frac{x_{it}}{\sqrt{T}}\right)^{j}\sum_{k=1}^{d}\tau_{0k}(t^{\widehat{\theta}_k}-t^{\theta_{0k}})$. We show that both contributions are $O_p(\ln T)$. By the mean-value theorem and Lemma 1(i),

$$
\begin{aligned}
&\left|\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left(\frac{t}{T}\right)^{\theta_{0i}}\sum_{k=1}^{d}\tau_{0k}(t^{\widehat{\theta}_k}-t^{\theta_{0k}})\right| \le \left|\frac{1}{\sqrt{T}}\sum_{k=1}^{d}\tau_{0k}\sum_{t=1}^{T}\left(\frac{t}{T}\right)^{\theta_{0i}}t^{\theta_{0k}}\left(t^{\widehat{\theta}_k-\theta_{0k}}-1\right)\right| \\
&\le C\sum_{k=1}^{d}|\tau_{0k}|\left|T^{\theta_{0k}+\frac{1}{2}}(\widehat{\theta}_k-\theta_{0k})\right|\frac{1}{T}\sum_{t=1}^{T}\left(\frac{t}{T}\right)^{\theta_{0i}+\theta_{0k}}\ln t \\
&\le C(\ln T)\sum_{k=1}^{d}|\tau_{0k}|\left|T^{\theta_{0k}+\frac{1}{2}}(\widehat{\theta}_k-\theta_{0k})\right|\left[\frac{1}{T}\sum_{t=1}^{T}\left(\frac{t}{T}\right)^{\theta_{0i}+\theta_{0k}}\right]=O_p(\ln T).
\end{aligned}
\tag{G.5}
$$

Similarly, from the mean-value theorem and Cauchy-Schwartz inequality, we see that

$$
\begin{aligned}
&\left|\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left(\frac{x_{it}}{\sqrt{T}}\right)^{j}\sum_{k=1}^{d}\tau_{0k}(t^{\widehat{\theta}_k}-t^{\theta_{0k}})\right| \le \left|\frac{1}{\sqrt{T}}\sum_{k=1}^{d}\tau_{0k}\sum_{t=1}^{T}\left(\frac{x_{it}}{\sqrt{T}}\right)^{j}t^{\theta_{0k}}\left(t^{\widehat{\theta}_k-\theta_{0k}}-1\right)\right| \\
&\le C\sum_{k=1}^{d}|\tau_{0k}|\left|T^{\theta_{0k}+\frac{1}{2}}(\widehat{\theta}_k-\theta_{0k})\right|\frac{1}{T}\sum_{t=1}^{T}\left|\frac{x_{it}}{\sqrt{T}}\right|^{j}\left(\frac{t}{T}\right)^{\theta_{0k}}\ln t \\
&\le C(\ln T)\sum_{k=1}^{d}|\tau_{0k}|\left|T^{\theta_{0k}+\frac{1}{2}}(\widehat{\theta}_k-\theta_{0k})\right|\sqrt{\frac{1}{T}\sum_{t=1}^{T}\left(\frac{x_{it}}{\sqrt{T}}\right)^{2j}}\sqrt{\frac{1}{T}\sum_{t=1}^{T}\left(\frac{t}{T}\right)^{2\theta_L}}.
\end{aligned}
\tag{G.6}
$$

From (G.5) and (G.6) we conclude that $\widetilde{\boldsymbol{D}}_{\boldsymbol{\theta}_0,T}^{-1}\sum_{t=1}^{T}\boldsymbol{z}_t(\boldsymbol{\theta}_0)\left(\boldsymbol{z}_t(\widehat{\boldsymbol{\theta}}_T)-\boldsymbol{z}_t(\boldsymbol{\theta}_0)\right)'\left[\begin{smallmatrix}\tau_0\\ \phi_0\end{smallmatrix}\right]=O_p(\ln T)$. ***(iv)*** Use $\boldsymbol{z}_t(\widehat{\boldsymbol{\theta}}_T)-\boldsymbol{z}_t(\boldsymbol{\theta}_0)=\left[\begin{smallmatrix}\boldsymbol{d}_t(\widehat{\boldsymbol{\theta}}_T)-\boldsymbol{d}_t(\boldsymbol{\theta}_0)\\ \boldsymbol{0}\end{smallmatrix}\right]$ to obtain $\widetilde{\boldsymbol{D}}_{\boldsymbol{\theta}_0,T}^{-1}\sum_{t=1}^{T}\left(\boldsymbol{z}_t(\widehat{\boldsymbol{\theta}}_T)-\boldsymbol{z}_t(\boldsymbol{\theta}_0)\right)\left(\boldsymbol{z}_t(\widehat{\boldsymbol{\theta}}_T)-\boldsymbol{z}_t(\boldsymbol{\theta}_0)\right)'\left[\begin{smallmatrix}\tau_0\\ \phi_0\end{smallmatrix}\right]$ $=\left[\begin{smallmatrix}\boldsymbol{D}_{d,T}(\boldsymbol{\theta}_0)^{-1}\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left(\boldsymbol{d}_t(\widehat{\boldsymbol{\theta}}_T)-\boldsymbol{d}_t(\boldsymbol{\theta}_0)\right)\left(\boldsymbol{d}_t(\widehat{\boldsymbol{\theta}}_T)-\boldsymbol{d}_t(\boldsymbol{\theta}_0)\right)'\tau_0\\ \boldsymbol{0}_{p\times 1}\end{smallmatrix}\right]$. For any $i\in\{1,2,\ldots,d\}$, the norm of the i^{th} component of the nonzero vector is

$$
\begin{aligned}
&\left|\sum_{k=1}^{d}\tau_{0k}\frac{1}{T^{\theta_{0i}+1/2}}\sum_{t=1}^{T}\left(t^{\widehat{\theta}_i}-t^{\theta_{0i}}\right)\left(t^{\widehat{\theta}_k}-t^{\theta_{0k}}\right)\right| \le \sum_{k=1}^{d}|\tau_{0k}|\frac{1}{T^{\theta_{0i}+1/2}}\sum_{t=1}^{T}t^{\theta_{0i}+\theta_{0k}}\left|t^{\widehat{\theta}_i-\theta_{0i}}-1\right|\left|t^{\widehat{\theta}_k-\theta_{0k}}-1\right| \\
&\le C\sum_{k=1}^{d}|\tau_{0k}|\left|\widehat{\theta}_i-\theta_{0i}\right|\left|\widehat{\theta}_k-\theta_{0k}\right|\frac{1}{T^{\theta_{0i}+1/2}}\sum_{t=1}^{T}t^{\theta_{0i}+\theta_{0k}}(\ln t)^2 \\
&\le C(\ln T)^2T^{-(\theta_L+\frac{1}{2})}\left|T^{\theta_{0i}+\frac{1}{2}}(\widehat{\theta}_i-\theta_{0i})\right|\sum_{k=1}^{d}|\tau_{0k}|\left|T^{\theta_{0k}+\frac{1}{2}}(\widehat{\theta}_k-\theta_{0k})\right|\left[\frac{1}{T}\sum_{t=1}^{T}\left(\frac{t}{T}\right)^{\theta_{0i}+\theta_{0k}}\right]=O_p\left(\frac{(\ln T)^2}{T^{\theta_L+\frac{1}{2}}}\right).
\end{aligned}
$$

(v) By similar steps as before, and invoking Theorem 2, it is easy to show that it suffices to bound $T^{-(\theta_{0i}+\frac{1}{2})}\sum_{t=1}^{T}(t^{\widehat{\theta}_i}-t^{\theta_{0i}})(u_t-\boldsymbol{\Omega}_{uv}\boldsymbol{\Omega}_{vv}^{-1}\boldsymbol{v}_t)$. Writing $u_t^{+}=u_t-\boldsymbol{\Omega}_{uv}\boldsymbol{\Omega}_{vv}^{-1}\boldsymbol{v}_t$ (as before), we have

$$
\begin{aligned}
&T^{-(\theta_{0i}+\frac{1}{2})}\sum_{t=1}^{T}(t^{\widehat{\theta}_i}-t^{\theta_{0i}})u_t^{+}=\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left(\frac{t}{T}\right)^{\theta_{0i}}(t^{\widehat{\theta}_i-\theta_{0i}}-1)u_t^{+}=(\widehat{\theta}_i-\theta_{0i})\frac{1}{\sqrt{T}}\sum_{t=1}^{T}(\ln t)\left(\frac{t}{T}\right)^{\theta_{0i}}u_t^{+}+o_p(1) \\
&=T^{-(\theta_{0i}+\frac{1}{2})}\left[T^{\theta_{0i}+\frac{1}{2}}(\widehat{\theta}_i-\theta_{0i})\right]\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left(\ln\frac{t}{T}\right)\left(\frac{t}{T}\right)^{\theta_{0i}}u_t^{+} \\
&\quad+T^{-(\theta_{0i}+\frac{1}{2})}\left[T^{\theta_{0i}+\frac{1}{2}}(\widehat{\theta}_i-\theta_{0i})\right](\ln T)\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left(\frac{t}{T}\right)^{\theta_{0i}}u_t^{+}+o_p(1)=\frac{1}{T^{\theta_{0i}+\frac{1}{2}}}O_p(1)+\frac{\ln T}{T^{\theta_{0i}+\frac{1}{2}}}O_p(1).
\end{aligned}
$$

This establishes (v). ∎

The currents upper bounds in the lemma above suggest that the RHS of (G.1) does not converge to a Gaussian mixture limiting distribution when $\boldsymbol{\theta}_0$ is unknown and has to be estimated. An additional simulation study was conducted to verify this claim. That is, we extend the simulation study on the Monte Carlo results for testing $H_0 : \phi_2 = 0$ versus $H_a : \phi_2 \neq 0$ to higher sample sizes. We consider serial correlation setting (D) and $\rho = 0.50$ as in the last column of Table 2. For sample sizes as large as 15000, the empirical size of feasible FMOLS estimator that relies on $\widehat{\theta}_T$ fluctuates around 11% (Figure 5). This indeed points towards a lack of asymptotic validity. On the contrary, FMOLS(θ_0) yields an empirical size close to 5%.

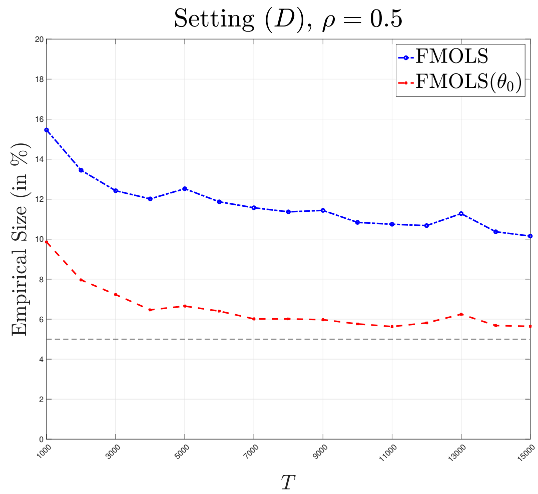


Figure 5: Empirical size of feasible and infeasible FMOLS estimators, see the note for Table 2.

H Further Empirical Results

H.1 Unit Root Tests

Table 6: The t-statistics for the ADF and DF-GLS unit root tests. The columns with header 'const' and 'const & trend' refer to the inclusion of only an intercept or both intercept and linear trend. Rejection of the unit root hypothesis at a 10% and 5% level are indicated with one and two stars, respectively.

	ADF				DF-GLS			
	const		const & trend		const		const & trend	
	GDP	CO_2	GDP	CO_2	GDP	CO_2	GDP	CO_2
Australia	0.287	-2.549	-2.050	-1.986	2.046	1.379	-1.577	-0.732
Austria	-0.055	-2.118	-1.943	-2.738	1.478	-1.143	-1.655	-2.718^{*}
Belgium	0.153	-2.336	-1.705	-2.818	2.041	-0.794	-1.287	-2.644
Canada	-0.575	-1.133	-2.020	-1.120	1.117	0.874	-1.894	-0.387
Denmark	-0.235	-2.446	-2.326	-0.136	1.393	0.410	-1.505	0.084
Finland	-0.362	-1.327	-2.315	-3.248^{*}	0.420	-0.076	-1.155	-3.217^{**}
France	-0.557	-2.438	-1.823	-1.858	1.087	-0.267	-1.470	-1.212
Germany	-0.374	-3.099^{**}	-2.767	-3.971^{**}	1.195	-0.726	-2.474	-2.080
Italy	-0.252	-1.546	-1.759	-1.987	1.213	0.354	-1.240	-1.860
Japan	0.010	-0.862	-1.733	-0.941	1.382	0.504	-1.272	-0.878
Netherlands	-0.106	-1.629	-2.247	-3.106	1.378	0.213	-1.679	-2.818^{*}
Norway	-0.680	-2.044	-2.064	-2.318	0.749	0.331	-1.017	-1.292
Portugal	-1.432	-0.455	-1.697	-1.676	-0.708	0.593	-0.741	-1.923
Spain	0.402	-1.243	-1.354	-1.994	1.487	0.959	-1.077	-2.014
Sweden	-0.789	-2.075	-2.289	-1.625	0.258	0.180	-1.513	-0.968
Switzerland	-1.093	-1.963	-2.785	-1.989	2.272	0.368	-2.447	-1.237
UK	-0.179	-0.721	-1.262	-0.402	2.446	-0.622	-0.608	-0.013
USA	-0.349	-2.055	-2.871	-1.322	2.409	-0.101	-2.708^{*}	-0.812

Note: Asterisks denote rejection of the null hypothesis at the ***1%, **5%, and *10% significance level.

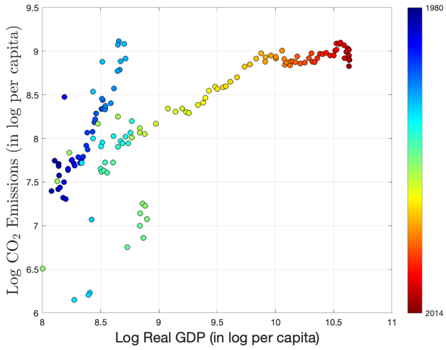
H.2 Perron and Yabu (2009) Test for Deterministic Trend Coefficient

The Perron and Yabu (2009) test is used to test for the presence of a deterministic trend function in the log per capita GDP series, see Table 7. The test allows for integrated or stationary errors. The details of the procedure can be found on page 61 of Perron and Yabu (2009). The asymptotic distribution of this test statistic is standard normal (quantiles are $z_{0.95} = 1.645$, $z_{0.975} = 1.96$, and $z_{0.995} = 2.58$).

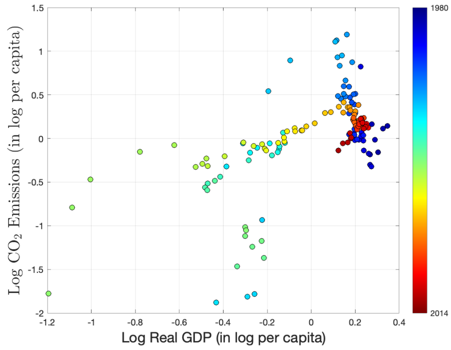
Table 7: Perron and Yabu (2009) test statistic for each of the 18 countries.

	$\widehat{PY}$		$\widehat{PY}$		$\widehat{PY}$
Australia	3.17	France	2.41	Portugal	2.16
Austria	2.19	Germany	1.91	Spain	2.31
Belgium	3.52	Italy	2.11	Sweden	7.12
Canada	3.33	Japan	2.93	Switzerland	3.91
Denmark	5.58	Netherlands	2.27	UK	3.60
Finland	4.27	Norway	5.85	USA	4.12

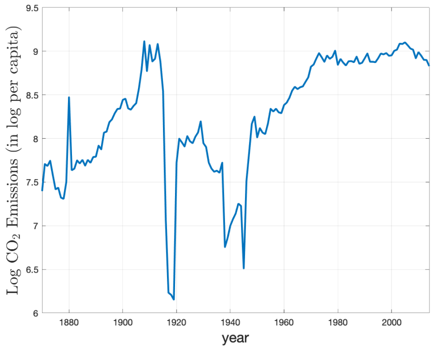
H.3 Overviews for Austria and Finland



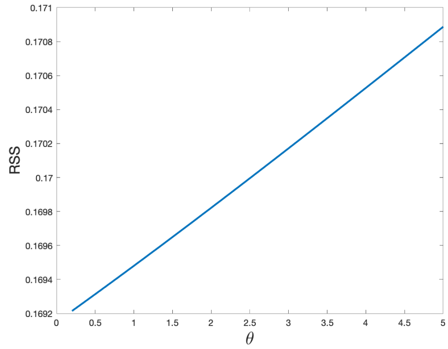
(a)



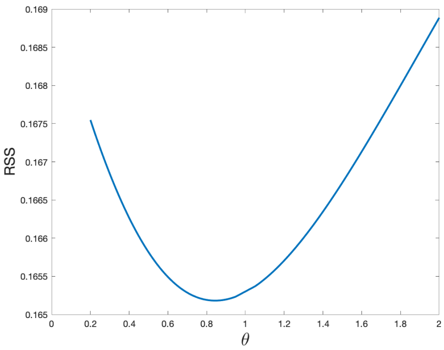
(b)



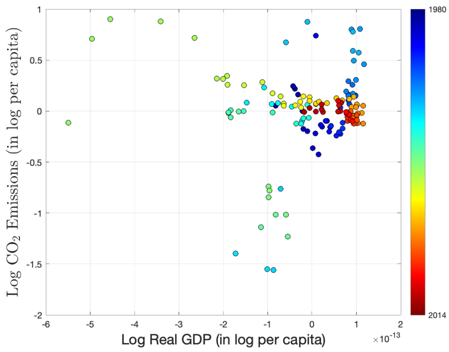
(c)



(d)



(e)



(f)

Figure 6: Overview graphs for Austria over 1870-2014. **(a)** log(GDP) versus log(CO_2) (both per capita). **(b)** As subfigure (a) but using detrended variables. **(c)** The log per capita CO_2 emissions time series for Austria. **(d)** The residual sum of squares (RSS) for the nonlinear model specification $y_t = \tau_1 + \tau_2 t + \phi_1 x_t + \phi_2 x_t^{\theta} + u_t$ for various values of θ. **(e)** The RSS as a function of θ for the flexible nonlinear trend specification $y_t = \tau_1 + \tau_2 t + \tau_3 t^{\theta} + \phi x_t + u_t$. **(f)** The relation between x_t and y_t after partialling out the constant, linear trend, and flexible deterministic trend.

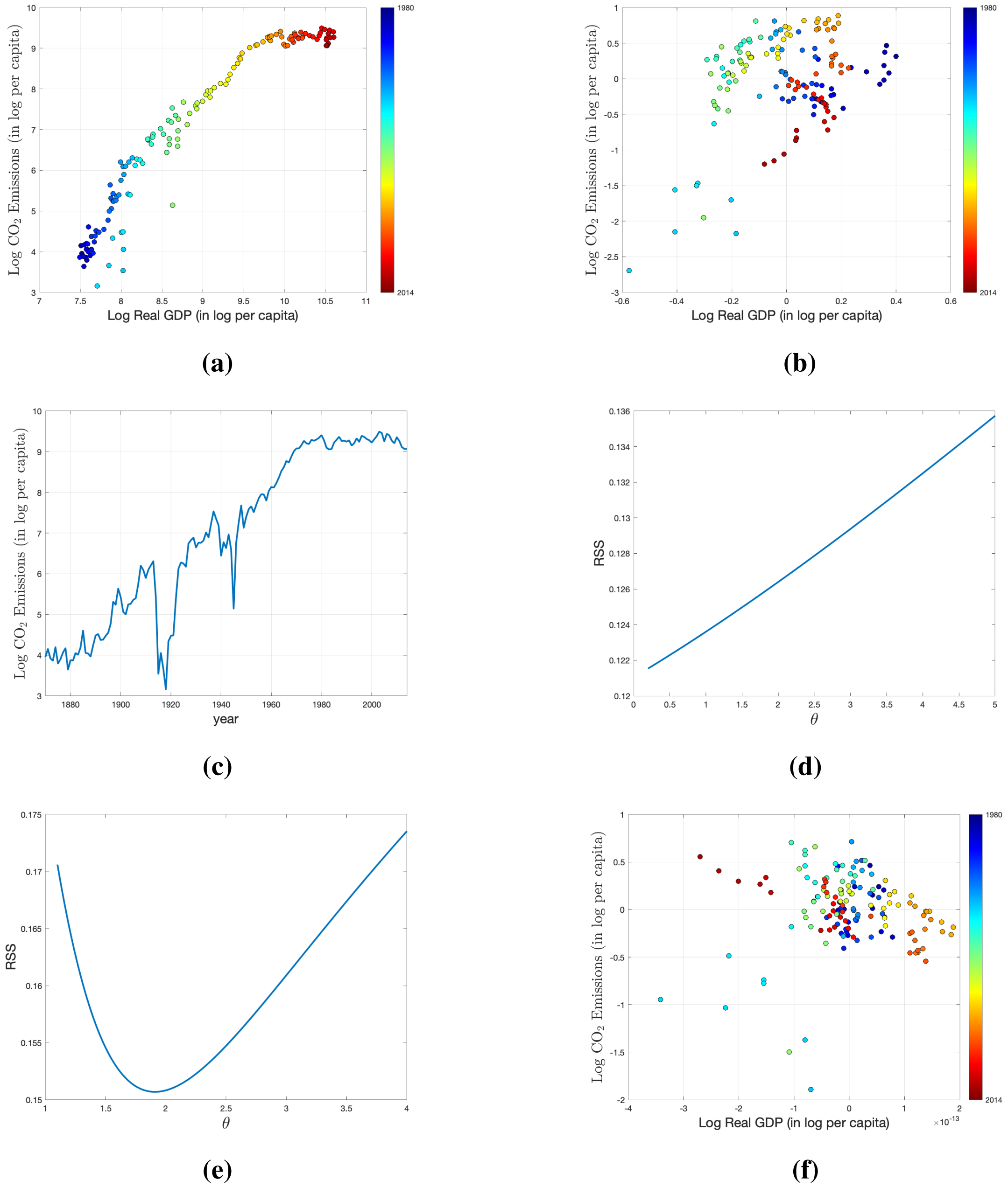


Figure 7: Overview graphs for Finland over 1870-2014. **(a)** log(GDP) versus log(CO_2) (both per capita). **(b)** As subfigure (a) but using detrended variables. **(c)** The log per capita CO_2 emissions time series for Finland. **(d)** The residual sum of squares (RSS) for the nonlinear model specification $y_t = \tau_1 + \tau_2 t + \phi_1 x_t + \phi_2 x_t^\theta + u_t$ for various values of θ. **(e)** The RSS as a function of θ for the flexible nonlinear trend specification $y_t = \tau_1 + \tau_2 t + \tau_3 t^\theta + \phi x_t + u_t$. **(f)** The relation between x_t and y_t after partialling out the constant, linear trend, and flexible deterministic trend.

H.4 Residual Series for Models (M1)-(M4)

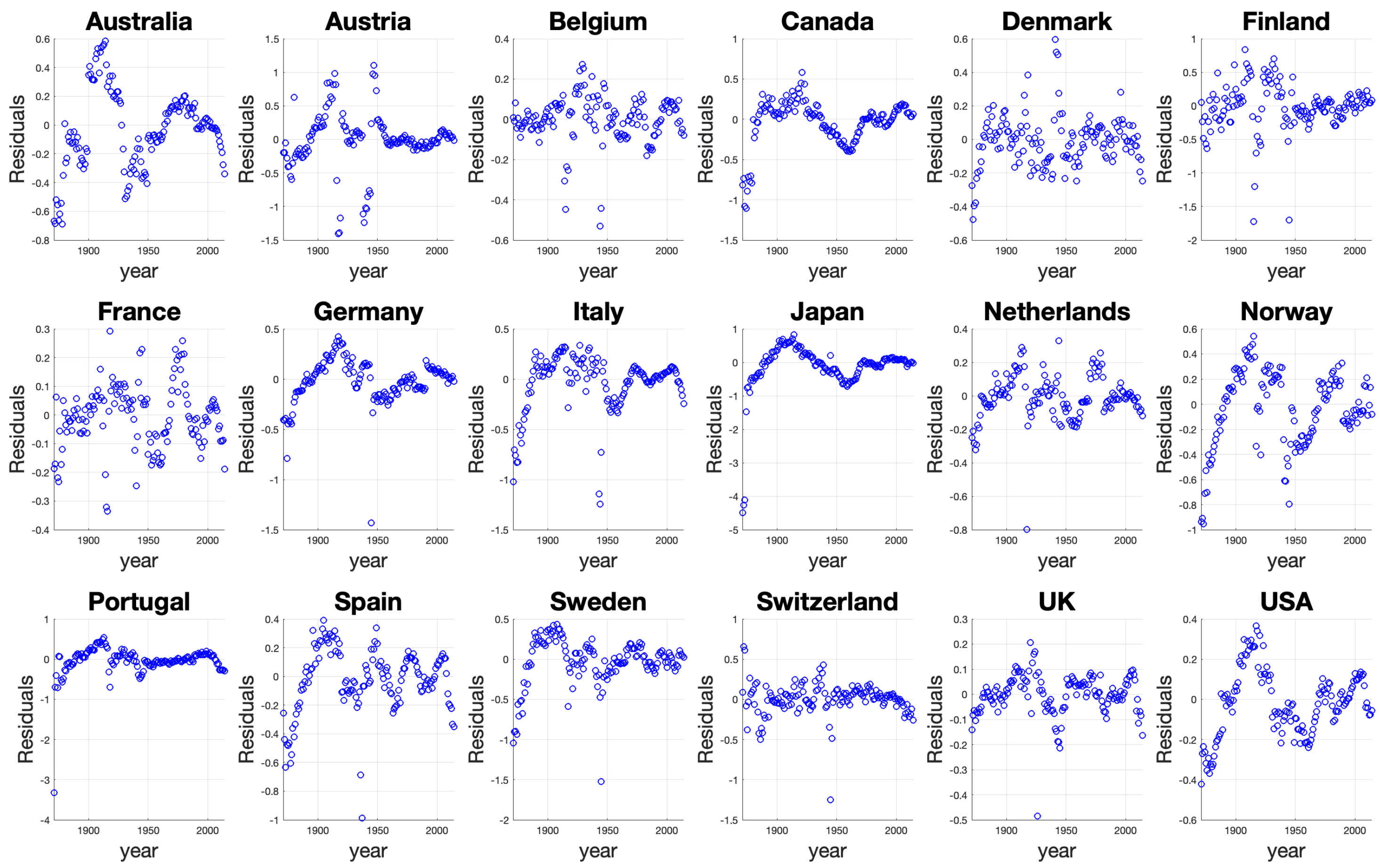


Figure 8: The residual series for each country under model specification (M1): $y_t = \tau_1 + \tau_2 t + \phi_1 x_t + \phi_2 x_t^2 + u_t$.

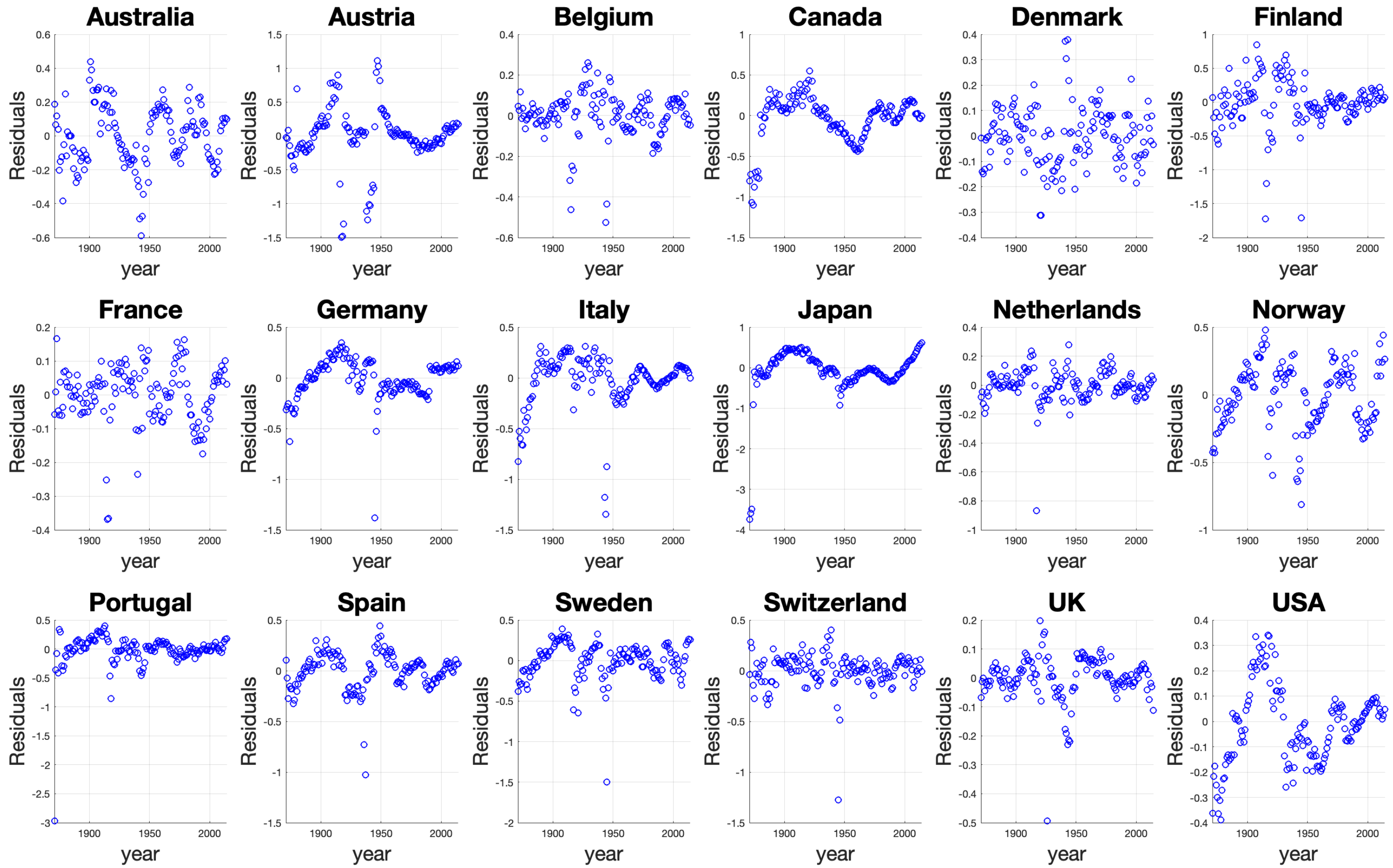


Figure 9: The residual series for each country under model specification (M2): $y_t = \tau_1 + \tau_2 t + \tau_3 t^2 + \phi_1 x_t + \phi_2 x_t^2 + u_t$.

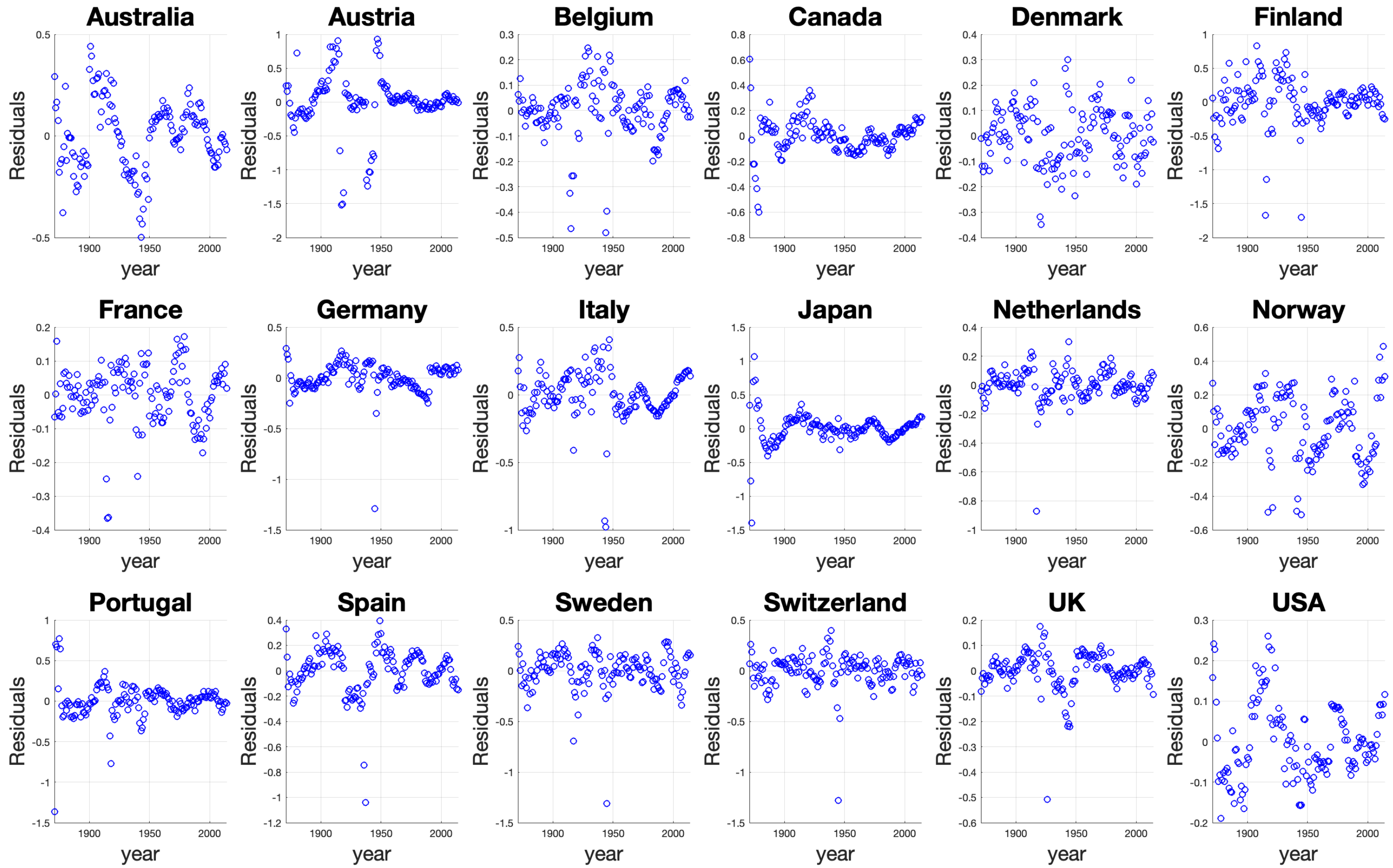


Figure 10: The residual series for each country under model specification (M3): $y_t = \tau_1 + \tau_2 t + \tau_3 t^{\theta} + \phi_1 x_t + \phi_2 x_t^2 + u_t$.

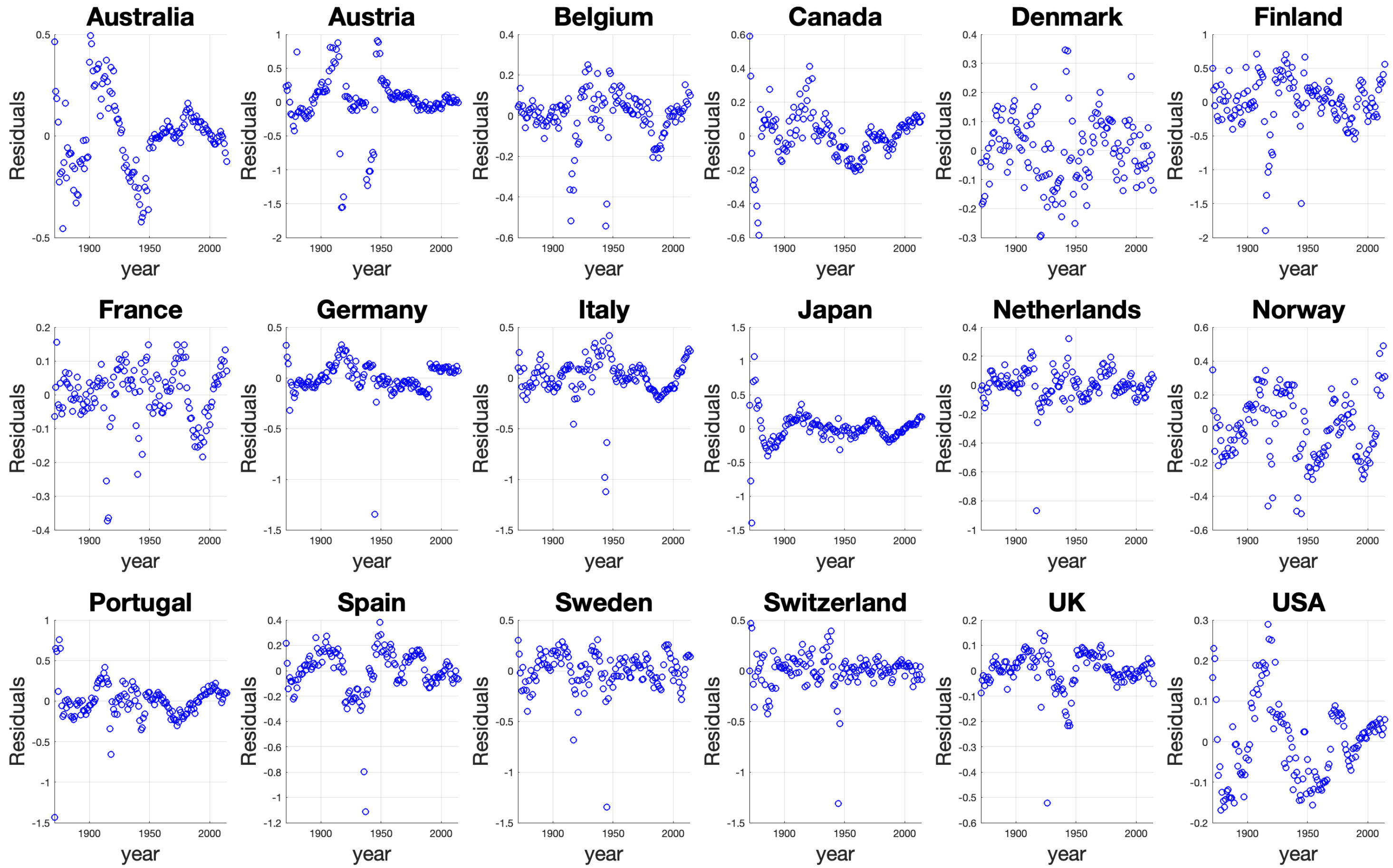


Figure 11: The residual series for each country under model specification (M4): $y_t = \tau_1 + \tau_2 t + \tau_3 t^{\theta} + \phi_1 x_t + u_t$.